The economics of diagnostic test: The cost-effectiveness of screening test for Gestational Diabetes Mellitus in Scotland

Noppcha Singweratham

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosoply (PhD)

Health Economics and Health Technology Assessment
Institute of Health & Wellbeing

University of Glasgow

January 2015

Abstract

Gestational diabetes mellitus (GDM) is the most common medical complication during pregnancy and is defined as carbohydrate intolerance with varying levels of severity, with the onset or first recognition occurring during pregnancy. A variety of different tests and guidelines have been used to screen for GDM over the past decade and due to this the prevalence's for GDM that are reported in studies tend to vary considerably. However, in Scotland there has been controversy over the method for the screening and the diagnosis of GDM, reflecting the lack of consensus for the diagnosis of this condition. This thesis therefore undertakes a cost-effectiveness analysis that compares four screening test strategies that use various combinations of screening and diagnostic tests with a strategy that involves no screening.

The first objective was to explore the economic approach to evaluating diagnostic testing in GDM. In consultation with experts and informed by comparable diagnostic testing models, the thesis adapted a conceptual model to the practice of GDM detection. The thesis found that the most appropriate model makes use of a combination of tests as either "negative dominant strategy" (NDS) or "positive dominant strategy" (PDS), allowing the clinician to consider test results in terms of the differences in false negative (FN) and false positive (FP) test results, as combining the tests in terms of NDS and PDS involves a tread-off between sensitivity and specificity.

A key input parameter of the model was identified to be the disease prevalence. However, due to limited known evidence in Scotland, the second objective was to assess the evidence on this key parameter. A systematic review was conducted to evaluate the prevalence of GDM, considering not just screening test characteristics, but also population characteristics. The review explores the possibility that variations in over half of the studies can be explained by ethnicity and diagnostic screening strategies and whether 75g or 100g oral glucose tolerance test (OGTT) methods of testing for GDM can cause variations in the results. Mothers with risk factors are prone to testing positive which in turn leads to higher prevalence rates compared to other populations and screening ethnic groups that have a high risk of developing GDM can indeed result in high prevalence estimates, ranging from 8.5% to 12.8%.

Decision making in healthcare over the past decade has increasingly been based on considerations of cost effectiveness, including national guidelines for GDM screening. The third objective was to summarize and appraise the present economic evaluation literature. Thus, a systematic review of economic evaluations, as outlined in the various cost and cost-effectiveness studies that have been published in recent

years, was performed to critically appraise the current analytical methods used to measure the cost-effectiveness of screening tests in order to develop a standardised economic model. Costs associated with the screening and management of GDM vary widely by country, ranging approximately from £2.42 to £50.9 per case detection, and are dependent on the tests used, the screening approach, how the costs are calculated and the prevalence of GDM in the population. The review of the CEA studies has significant implications for future research and policy making and as such long term consequences are appropriate outcomes for the CEA of screening tests for GDM in order to capture all long term adverse health outcomes for GDM. Therefore, the economic evaluation for GDM should account for the effectiveness of postpartum screening for type 2 DM.

The aforementioned conceptual model and the results of the systematic reviews of diabetes prevalence and GDM screening cost-effectiveness have been synthesised into a model to estimate the cost-effectiveness of screening for GDM based on four screening guidelines that include 1) SIGN 2001 (random plasma glucose followed by 75g OGTT) , 2) NICE 2008 (risk factors screening followed by fasting plasma glucose and 75g OGTT) , 3) Consensus 2010 (75g OGTT) and 4) SIGN 2010 (risk factors screening followed by 75g OGTT) versus 5) no screening. This probabilistic model was used to estimate and compare the costs and quality adjusted life years (QALYs) of screening tests for GDM. Three independent decision trees, for case detections, short term complications in the first year and long term complications over the lifetime, explored and considered the combinations of screening and diagnostic tests in terms of NDS and PDS. Independent decision trees would allow policy makers to focus on each part of the model separately. The primary outcomes of the analysis were the incremental cost per case identification, one year QALYs for short term complications and lifetime QALYs for type 2 diabetes mellitus for long term complications. Case identification was insufficient for policy makers because it fails to take account of the consequence of false positive and false negative results of tests.

For short term complications, the incremental cost-effectiveness ratio was £46,760 per QALY for the two-step approach with SIGN 2001 (NDS) using 75g OGTT to confirm any positive random plasma glucose (RPG) before treatment compared with no screening. At willingness to pay £30,000/QALY, this strategy has 64% probability of being cost-effective for short term complications. The cost effectiveness of screening tests for GDM, to prevent short term complications, is dependent on the probability of GDM being undiagnosed. Additionally, treatments during gestation are important as they reduce additional costs that may be required to treat serious adverse complications.

PDS screening where all pregnant women with one or more high risk factors are requested to undertake 75g OGTT diagnostic test, proposed by SIGN 2010, is the most cost-effective strategy in long term complications. SIGN 2010 (PDS) has higher QALY (80.9736) and is less expensive (£4,088) than the other strategies and dominates the other screening test strategies for long term complications. At a threshold of £30,000/QALY, the CEA illustrates that the probability that SIGN 2010 (PDS) will be cost-effective is approximately 55.8%. The cost effectiveness of screening tests for GDM, to prevent long term complications, is dependent on the probability of GDM being over diagnosed. If mothers have received previous diagnoses of GDM, this should trigger regular screening for type 2 DM so that it is discovered early on, before the onset of symptoms or the development of complications associated with type 2 DM. Postpartum screening and the subsequent treatment of GDM presents an important opportunity to reduce type 2 DM.

This thesis provides the first economic model of a screening test using independent decision trees, split by NDS and PDS. By using NDS and PDS, decision makers can interpret the combination of test results. This better presents the consequences of false positive and false negatives and a trade-off between sensitive and specificity. The thesis finds novel value of applying this methodology to GDM screening. By using independent decision trees for NDS and PDS, the model was able to identify long term complications as the most important factors affecting the results of screening test strategies. Currently, all guidelines for GDM screening tests which also included two Scottish guidelines SIGN 2001 and SIGN 2010 are performed as NDS with regard to both screening tests and postpartum screening. Thus, in Scotland, policy makers or clinicians should consider SIGN 2010 with PDS for screening tests for GDM in order to prevent long term complications. Lastly, this work is also the first to apply the expected value of information (EVPI) to the area of GDM screening. The population EVPPI of £784,042 for long term complications shows that there is greater uncertainty with respect to long term complications and that collecting information on long term complications is likely to be worthwhile. Thus, it captures the long time horizons in screening programmes required for decisions about the value of further research and the expected payoff of conducting further research to resolve the model uncertainties.

Table of contents

List of tables

List of figures

Presentations and working papers

The presentations and working paper are a result of the research conducted for this thesis.

Conference presentation

Conference Abstract: presentation, The economics of diagnostic testing: The cost-effectiveness of screening tests for gestational diabetes mellitus in Scotland. 14th Biennial SMDM (Society for Medical Decision Making) European Meeting Olso, Norway,June 10-12,2012

Conference Abstract: presentation, Screening test for Gestational diabetes mellitus: A systematic review of the economic evaluation and modelling. Institute of Health and Wellbeing Student-led Conference, University of Glasgow, November 26th, 2012, Glasgow, Scotland

Conference Abstract: poster presentation, Prevalence of gestational diabetes mellitus: A systematic review of population-base studies. 7th International Diabetes, Hypertension, Metabolic Syndrome & Pregnancy (DIP) Symposium. Florence, Italy March 13-16, 2013

Acknowledgement

First and foremost I would like to express my sincere gratitude to my supervisor Professor Andrew Briggs for his continuous support and guidance throughout my work as well as his incredible knowledge of the field. His patience, motivation and enthusiasm have been an invaluable help in completing my thesis.

I am also extremely grateful to Professor Elisabeth Fenwick who was always ready to help whenever it was needed during my time at Glasgow University, although not required to do so. Your help is very much appreciated.

I also wish to acknowledge the funding sources that made my Ph.D. work possible as well as the great support they offered me during my time in the UK. I was funded and supported by Praboromarajchanok Institute, the Thai Ministry of Public Health and the Thai Royal Government.

I would also like to take this opportunity to thank my various colleagues at Glasgow University who have provided tireless encouragement and support throughout the process. In particular I would like to mention Dr Claudia Geue and Dr Kathleen Anne Boyd. I thank them both for their encouragement and optimism.

To my friends that have offered advice and encouragement, I am deeply applicative for all that they have done. To Colin Rogers for contributing his time to support me when most needed and to Seamus Kent for his insightful discussions and opinions that have helped so much.

Last but not least I must express my appreciation to my family for their love and faithful support for believing in me and my abilities over the past few years.

Author’s declaration

I declare that, except where reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name: Noppcha Singweratham

Definitions/abbreviations

ABCD	Association of British Clinical Dialectologists
AC	Abdominal circumference
ACER	Average cost effectiveness ratios
ACHOIS	Australian Carbohydrate Intolerance Study in Pregnant Women
ACOG	American College of Obstetricians and Gynecologists
ADA	American Diabetes Association
ADIPS	Australian Diabetes in Pregnancy Society
AUC	Area under the curve
BMI	Body mass index
BPI	Brachial plexus injury
C&C	Carpenter and Coustan
CA	Cost analysis
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CD	Case detection
CDA	Canadian Diabetes Association
CE Plane	Cost-effectiveness plane
CEA	Cost-effectiveness analysis
CEAC	Cost effectiveness acceptability curve
CER	Cost-effectiveness ratios
CI	Confidence interval
CMA	Cost-minimization analysis
CN	Chance node
CP	Case prevention and adverse complication
CREST	Clinical Resource Efficiency Support Team
CUA	Cost-utility Analysis
DALY	Disability-adjusted life-years
DES	Discrete event simulation
DM	Diabetes mellitus
DN	Decision node
DRG	Disease related groups
DTT	Diagnostic test threshold
EASD	European Association for the Study of Diabetes
ECDC	Expert Committee on the Diagnosis and Classification of DM
EORT	European organisation for research and treatment of Cancer
EQ-5D	EuroQol 5D
EVDI	Expected value of diagnostic information
EVI	Expected value of information
EVPI	Expected value of perfect information
EVPPI	Expected value of perfected information for a parameter
EVSI	Expected value of sample information
FACT	Functional Assessment of Cancer Therapy scale
FBG	Fasting blood glucose
Fetal AC	Fetal abdominal circumference
FLIC	Functional living index cancer
FN	False negative
FNR	False negative rate
FP	False positive

FPF	False positive fraction
FPG	Fasting Plasma Glucose
FPR	False positive rate
GCEA	Generalized cost-effectiveness analysis
GCT	Glucose challenge test
GDM	Gestational diabetes mellitus
HbA1c	Glycosylated hemoglobin A1c
HCHS	Hospital and Community Health Services
HDL	High density lipoprotein
HOQoL	Health related quality of life year
HTA	Health technology assessment
HUI	Health Utility index
ICG	International Consensus Guideline
IADPSG	International Association of Diabetes and Pregnancy Study Group
IC	Incremental cost
ICER	Incremental cost effectiveness ratio
IGT	Impaired glucose tolerance
INMB	Incremental net monetary benefit
LAC	Long term adverse complications
LE	Life expectancy
LR	Likelihood ratio
MFMU	Maternal-Fetal Medicine Units network
NDDG	National Diabetes Data Group
NDS	Negative dominant strategy
NHS	National Health Service
NHS EED	NHS Economic Evaluation Database
NICE	Institute for Health and Clinical Excellence
NICU	Neonatal Intensive care unit
NMA	Network meta-analysis
NMB	Net monetary benefit
NPV	Negative predictive values
OGTT	Oral glucose tolerance test
OSA	One way sensitivity analysis
PAEs	Perinatal adverse events
PDS	Positive dominant strategy
PPV	Positive predictive value
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
QLC-C30	Quality of Life Core 30
RBG	Random blood glucose
RCT	Randomized control trial
ROC	Receiver operating characteristic
RPG	Random Plasma Glucose
RSCL	Rotterdam Symptom Checklist
SA	Sensitivity analysis
SC	Short term complications
SD	Standard deviation
SE	Standard error
SF-12	Short Form 12
SF-36	Short Form 36
SF-6D	Short Form 6D

SG	Stand gamble
SHPIC	Scottish Health Purchasing Information Center
SIGN	Scottish Intercollegiate Guidelines Network
SMC	the Scottish Medicine Consortium
SOGC	Society of Obstetricians and Gynaecologists of Canada
SPC	Serious perinatal complication
STD	Signal detection theory
TN	True negative
TNR	True negative rate
TP	True positive
TPR	True positive rate
TTO	Time treat-off
UKPDS	UK Prospective Diabetes Study
VAS	Visual analogues scale
VOCI	Value of clinical information
VOI	Value of Information
WHO	World Health Organisation

Chapter 1 Introduction

1.1 Introduction

Increasing costs of health care relative to other sectors, due to the labour intensive nature of health care provision, has lead to pressures on health care budgets. Moreover, health care resources are limited by total funds available, as well as through competition with other areas, such as housing, education and finance. This raises the question of how to make decisions about the proper allocation of funds (Hine, 1999). Nowadays, governments in many countries are faced with rising health care costs due to rapidly growing populations, increasing demands for health care services, and the introduction of new health technology. As a result, governments are under pressure to justify their resource allocation. Many countries around the world apply economic evaluation to the decision-making process for health, including Australia, Canada, the United States of America and many European countries. Each year more than £100 billion is spent by the UK National Health Service (NHS) on health care in the UK (Department of Health, 2010). With new technologies in health care continually being discovered, there is a need for decision making strategies and prioritization of health care when faced with limited budgets. Rising expenditure in health care is often associated with new technologies and health care programmes (Cunningham, 2001). Given the limitations in health care resources, there is increased interest in assessing the value for money, or economic efficiency of health care programmes. This is achieved through economic evaluation, where the costs and consequences of alternative intervention strategies are compared (Drummond, 2005b). Therefore, economic evaluation has become part of modern health care evaluation.

Diagnostic tests can monitor the benefits and/or side effects of appropriate treatment. An error in diagnostic test results may result in over or mis-diagnosis, also referred to as false positive and false negative results, respectively. Over diagnosis refers to the situation where screening programmes detect disease which is not in fact present (Moynihan et al., 2012). People without disease symptoms can then be diagnosed and treated for a disease which may in turn lead to unnecessary and possibly harmful treatment, and resulting in disutility for these patients. On the other hand, miss diagnosis refers to patients that have disease but test negative during screening which may delay access to beneficial treatment. Diagnostic tests come with costs and consequences both of which are appropriate subjects for economic analysis. Because of increasing health care costs and limited resources, obstetricians must pay attention to the cost-effectiveness of the diagnosis and treatment programmes. In addition, obstetricians should be concerned with several important factors in the cost analysis of

health care. Diagnostics consist of using the tests to categorise the population into two groups or help physicians revise disease probability for their patients. A test may be performed to prove that a person is free from disease. Additionally, the aim of all diagnostic testing is to refine this probability to the point where the clinicians can confidently assign a treatment to the correct patient. Each diagnostic test, whether it is a symptom, sign, laboratory or radiological examination, results in a change in the clinician's probability of disease in their patients.

Gestational diabetes mellitus (GDM) is the most common medical complication of the pregnancy period. GDM is defined as carbohydrate intolerance of variable severity, with onset or first recognition during pregnancy, whether or not insulin is used, and regardless of whether diabetes persists after pregnancy (WHO, 1999a). The prevalence of women with GDM has increased over time, impacting between 1% - 16% of women, depending on the diagnostic criteria and population studies (King, 1998). There are three main reasons for screening for GDM. First of all, GDM is a significant complication and metabolic disorder of pregnancy, and disappears after delivery (Koukkou et al., 1995) (Carpenter and Coustan, 1982). Secondly, GDM is associated with adverse effects for both mother and child. These include fetal macrosomia and perinatal mortality in children, hypertension disorders, pre-eclampsia, and cesarean section in mothers, and future diabetes in both (Persson and Hanson, 1998) (Greco et al., 1994). Thirdly, identifying this group of women can not only prevent perinatal morbidity but also reduce risks to health through diet and lifestyle management, oral medication programmes and insulin therapy (Jovanovic, 1998).

Screening tests for GDM are performed by either selective or universal screening. Normally, screening and diagnostic tests are performed between 24 and 28 weeks gestation because the diabetogenic effect takes place at this time in pregnancy (Naylor et al., 1997) and there is sufficient time during pregnancy to start treatment or therapy programmes (Greene, 1997). Screening tests are a controversial topic in terms of the appropriate screening approach of pregnant women with GDM (Greene, 1997) (Jarrett et al., 1997) (Soares J de AC, 1997) and screening test strategies for GDM are still debated with no consensus having been established yet (Metzger and Coustan, 1998). These strategies are widespread screening of the population, the timing of screening, screening test techniques and the threshold of diagnosis. Moreover, in screening for GDM there is also controversy regarding the use of universal or selective screening (Vogel et al., 2000). The Australasian Diabetes in Pregnancy Society (ADIPS) Management Guidelines for GDM strongly recommends that screening tests for GDM should be used to screen all pregnant women. If the resources are restricted, high risk factor groups will be screened (Hoffman et al., 1998). Common risk factors for GDM

include advanced maternal age (women > 40 years), a family history of diabetes, obesity, previous adverse pregnancy outcome, non-white ethic origin and being a current smoker (Santini and Ales, 1990) (Sacks et al., 1987). However, testing by universal screening, which tests every woman in every pregnancy, comes at a cost, and in addition those women who are diagnosed with GDM will use further health care resources, with cost implications.

In terms of the management of GDM, effective treatment demonstrates that diagnosis and treatment of GDM results in a reduction in adverse effects for both mother and child. Two randomised control trials have shown that interventions in women diagnosed with GDM using dietary advice, monitoring and management of blood glucose, are effective in reducing birth weight, the rate of growth for gestational age infants, and perinatal morbidity. Treatment of GDM reduces serious perinatal complications and may also improve women's health-related quality of life during pregnancy and after delivery. Two studies found that treatment of women with GDM, including dietary advice, blood glucose monitoring, and insulin therapy, reduced the rate of serious adverse outcomes (defined by one or more of the following: death, shoulder dystocia, bone fracture, and nerve palsy) from 4 percent to 1 percent (Crowther et al., 2005) (Landon et al., 2009)

In the past, convenient screening tests such as glycosuria, the measurement of blood Glycosylated hemoglobin A1c (HbA1c) and random plasma glucose tests (RPG) were used to classify GDM, which were not adequately sensitive to measure glucose levels. At present, the test most used internationally is a 2h 75g oral glucose tolerance test (OGTT) and 3h 100g oral glucose tolerance test (OGTT). The 2h 75g OGTT is recommended by the World Health Organisation (WHO) and is widely used outside the United States of America, especially in Europe and Japan (Alberti and Zimmet, 1998). On the other hand, The American Diabetic Association (ADA) recommends screening for gestational diabetes with the 3-hour 100g oral glucose tolerance test, it is commonly used in the United States of America (Martine et al., 2007). However, in the UK, two national surveys of UK obstetric units have shown a lack of consensus on screen tests for GDM, with some health care centres using the ADA guidelines and others implementing the WHO guidelines. Moreover, screening and diagnostic tests, cut-off values, timings, and subsequent management programmes vary widely in the UK (Mires et al., 1999) (Hanna et al., 2008). The national survey of the Association of British Clinical Diabetologists (ABCD) reported 82 percent of centres in the UK providing routine screening for GDM, half of those centres using universal screening and half selective screening. This survey reports a widely varied prevalence of GDM from 0.1-10% (Median 1.5%) (Hanna et al., 2008).

In Scotland, there has been controversy over the best method of screening and specific diagnostic criteria for the detection and definition of GDM, with no accepted international consensus to this day. The Scottish Intercollegiate Guidelines Network (SIGN) published two guidelines in 2001 and 2010 called “Management of Diabetes - A National Clinical Guideline” each with different screening test methods (SIGN, 2001) (SIGN, 2010). There is very limited information available about the cost of screening and diagnostic tests for GDM, and the utilisation of resources involved. However, it is commonly accepted that detection of GDM is of great importance in implementing a proper strategy for reducing maternal and perinatal complications. With the limitations in health care and budgets, economic evaluation can help to determine the most cost-effective screening and diagnostic tests among numerous intervention options available. This evidence can help policy decision-makers set priorities in terms of budget allocation (Singh et al., 2001). The decision analysis method has been applied to evaluate screening and diagnostic tests in terms of possible clinical consequences. There is no study, however, on the economic evaluation of screening and diagnostic tests for GDM in Scotland. This study compares different strategies, of screening for GDM in Scotland.

1.2 Objectives of this thesis

The overall aim of this thesis is to study the cost-effectiveness of screening and diagnostic tests for GDM in Scotland. To estimate GDM detection rates and the proportion of women identified for subsequent testing involves modelling a relationship between prevalence and the positive and negative predictive value of testing. This study aims to inform policy makers on this important issue by establishing the optimal mix of cost-effective screening and diagnostic tests of GDM. This study compares four strategies of screening for GDM against no screening in different organisations in Scotland.

1.2.1 Objective 1: To explore the economic approach to evaluating diagnostic tests

The first objective is to describe the combination of the test approaches in evaluating diagnostic tests. The combination of the tests as either “negative rule” or “negative dominant strategy” and either “positive rule” or “positive dominant strategy” are described and illustrated. Positive and negative dominant strategies may help the clinician to interpret the combination of test results and illustrate this new technique for economic test evaluation.

1.2.2 Objective 2: To obtain disease prevalence, diagnostic test accuracy and health benefits of outcomes of GDM by systematic review

The second objective is to ascertain data for the input parameters by means of a systematic review. All economic evaluation studies are obliged to include all parameters in the economic model. This study also requires many parameters to populate the model, such as GDM prevalence, sensitivity and specificity of the screening test and proportion of adverse outcomes in both mother and offspring. This paper includes a systematic review to reassess the prevalence and incidence of gestational diabetes mellitus in the general pregnancy population.

1.2.3 Objective 3: To review the economic evaluation of screening tests for GDM

Screening tests for GDM are widely reported to be cost-effective interventions. Therefore, the third objective aims to review systematically and critically an economic evaluation of screening tests for GDM. The review assesses the evidence for the cost-effectiveness of different approaches to screening tests, taking into account the appropriateness of the model, and assesses the data requirements used to derive cost and effectiveness of screening tests for GDM. The findings for this review should help to provide more reliable information about the economic evaluation of screening tests of GDM and develop a new and appropriate model for GDM screening tests.

1.2.4 Objective 4: Estimate the cost-effectiveness of screening test for GDM

The fourth objective is to compare the cost-effectiveness of both universal and selective screening tests for GDM in Scotland. A basic decision tree structure is used in developing the model, based on five strategies of both universal and selective screening used to identify glucose intolerance during pregnancy. This part of the analysis focuses on the health outcomes of GDM. Quality adjusted life years (QALY) in both mothers and offspring are split into short and long term complications. The most appropriate assessment for effective comparison across disease states include health-related quality of life year (HRQoL) measures such as EuroQol and the 5 dimensional (EQ-5D), which can be employed to derive QALY. This paper will assume a willingness to pay £30,000 per QALY.

1.2.5 Objective 5: Impact of uncertainty

This objective will address the issues of impact of uncertainty during every stage of the economic evaluation of GDM. Probabilistic sensitivity analysis is used to estimate the probability that each strategy will be cost-effective at a given willingness to pay for each additional unit of benefit.

1.3 Structure of the thesis

The structure of this thesis is as follows:

Chapter 2 summarises a narrative of the grounds for economic evaluation of health care intervention. This chapter introduces two important frameworks of economic evaluation and decision analytic modelling. In the first framework, the key concepts for economic analysis in health care are introduced before exploring the health economic evaluation technique. The chapter briefly describes a general overview of health economic evaluation; perspectives and comparators of study; costing and time horizons are discussed. The second framework is based on decision modelling for health economic evaluation, which describes cost-effectiveness analysis based on cohort modelling; such as decision tree modelling and Markov modelling. The finer details of the rules governing cost-effectiveness analysis are also examined.

Chapter 3 mainly describes the fundamental terminology of diagnostic test evaluation to support the methodology of this thesis. In order to present the diseases present and absent in the population, the accuracy of the test are also introduced. This chapter will focus on Bayesian inference by giving an introduction to the theorem and demonstrating an example of how positive and negative predictive values from sensitivity and specificity are calculated and to show how to apply this technique to the economic evaluation of diagnostic tests. The two component tests under either the positive dominant strategy or the negative dominant strategy are discussed.

Chapter 4 provides a background review of the literature regarding GDM by means of a systematic review. This chapter begins by outlining the natural history of GDM and the treatment for the disease in order to start the conceptualisation around and economic evaluation for GDM. Generally, before developing an economic evaluation, it is essential to fully understand the nature of the disease under scrutiny, including how the disease presents itself, how the disease propagates and how treatment affects the disease. The short and long term complications that occur in both mothers and offspring are briefly reviewed. The testing techniques, the various screening and diagnostic tests in current

use are also discussed. The systematic review of GDM will provide all relevant data for the input parameters to populate the economic model for GDM screening tests.

Chapter 5 conducts a comprehensive systematic review of the prevalence of GDM. This chapter reviews an article which reports on the prevalence of GDM in the general population. The selected articles are presented and prevalence is estimated. The prevalence estimates will be discussed and synthesised in this chapter, and are used as input parameters in the model. This addresses objective two.

Chapter 6 presents and critiques a review of the relevant published literature on cost analysis and economic evaluation of GDM screening by means of a systematic review. The key concepts of economic evaluation identified in chapter 2 such as perspective, patient group, comparators, outcome measures, extrapolation, and the results of the evaluations and the handling of uncertainty are criticised in this particular paper. Following this is a comparison of all papers discussed in which important methodological issues and gaps are highlighted. The result of this review provides a new economic evaluation model of GDM in Scotland. This chapter supports objective 3.

Chapter 7 aims to construct a model of cost-effectiveness for GDM by integrating the natural history of GDM, diagnostic test evaluation and health economic evaluation. This chapter assesses the relative costs and effectiveness of single outcomes (case identification). The two components of the tests from chapter 3 are applied to construct the decision tree model. The results from the cost-effectiveness analysis of case identification are then presented.

Chapter 8 develops and presents the results of a decision analysis model of GDM screening tests in terms of short and long term adverse complications. The developments of the decision tree models in both short and long term adverse complications are detailed. The outcome measures of effectiveness were QALY in both short and long term complications. In order to determine QALY in both mothers and offspring in short term complications, this chapter calculates the QALY over 1 year from health-related quality of life scores. With respect to long term adverse complications in both mothers and offspring with diabetes mellitus life tables, life expectancy and quality-adjusted life expectancy are calculated. This addresses objective four.

Chapter 9 conducts a cost-effectiveness analysis to compare four screening test strategies and a strategy that involves no screening test, following the development of the decision tree in chapters 6 and 7. This chapter explores uncertainty in the model result and demonstrates that using a cost-effective acceptability curve, where results

can be examined in terms of decision uncertainty, is recommended. The Value of Information (VOI) technique is presented to give meaningful recommendations to funding organisations and decision making bodies. VOI can eliminate the possibility of making the wrong decision and can also present the expected opportunity loss surrounding the decisions.

Chapter 10 summarises and discusses the main findings from analyses in the empirical chapters. Beside this, recommendations based on the findings are also provided, in order to demonstrate the knowledge gained from the literature and past research. This chapter also explains how the overall aim of the project was accomplished through the work undertaken. Limitations of the thesis are discussed and the scope for future research is carried out.

Chapter 2 Economic evaluation and health decision making

2.1 Introduction

This chapter aims to explore the grounds for economic evaluation of health care interventions. This chapter is split into two halves; the economic evaluation framework and the decision analysis modelling framework.

The first section of the first half discusses an eight step framework for conducting economic evaluations. This section introduces the key concepts for economic analysis in health care before exploring health economic evaluation techniques. The subsequent section further describes the components of health economic evaluation, including perspectives and comparators of study, costs and the time horizon. Then, the different types of outcome measures are followed by a discussion of the time horizon, discount rate and sensitivity analysis. Following this, descriptions of the various models of economic evaluation techniques that are commonly employed are discussed. It describes the cohort model, the main type of decision model used in the field.

The second half of the chapter begins by discussing the six steps of the framework for decision analytical modelling. Firstly the decision problem is outlined and then the decision model itself is structured before identifying and synthesising the available evidence. The systematic review and meta-analysis in this section are discussed in terms of how to access appropriate information for using an input parameter in an economic model. The following section examines the structure of the model for decision-making in detail, the deciding factors and cost-effectiveness analysis. Cost-effectiveness analysis and the deciding factors are described to decision makers using methods such as incremental cost-effectiveness ratio, net benefit and net monetary benefit. Following this, there is a discussion of the methods for handling uncertainty and heterogeneity in models for decision-making, in particular the use of probabilistic sensitivity analysis to reflect parameter uncertainties. Then, the value of information methods are explored and used to look at how future research studies may be efficiently designed.

2.2 Economic evaluation

In 1948, the National Health Service (NHS) published three core principles: the NHS would aim to meet the needs of everyone, be free at the point of delivery and be based on clinical need, not ability to pay (National Health Service, 1948). Currently, those three components are being applied. Difficult decisions are constantly being made because resources are limited, and include such issues as which new treatments and procedures to finance (and for whom) and which older treatments should be replaced.

Academic researchers and others, especially economists, can support and guide the policy maker with these difficult decisions by conducting health economic evaluations on the value of pharmaceutical products, devices and other interventions. The value of one treatment in terms of its costs and effects is compared to the costs and effects of other treatments, currently in practice. The main purpose of providing information to health care decision makers is to allocate resources efficiently in terms of costs and quality of life, within the health care sector (Evans et al., 2004). Economists attempt to make explicit one set of criteria that can be used to decide between different uses of resources in short-supply (Drummond, 2005b). Thus the choices of which treatment or intervention should be accepted and/or eliminated are made based on economic value and health related quality adjusted life years.

Nowadays, many countries around the world incorporate economic evidence into the decision making process for health and reimbursement decisions in health care. This includes many European countries, Australia and Canada (CADTH, 2006) (Andronis et al., 2009) (NICE, 2013). For example, NICE is an organisation independent from the government, which is the relevant decision making agency in England & Wales, responsible for providing guidance on the use of health technologies and the implementation of public health programmes. In 2008, NICE proposed guidelines for methods of technology appraisal (NICE, 2008c). These guidelines were revised and proposed in a new version in 2013 (NICE, 2013). In both versions cost- effectiveness analysis (CEA) is recommended as a highly integrated component of the technology appraisal process by NICE. NICE brings together all evidence in order to decide whether the adoption of a technology (treatment or drug) represents good value for the NHS during the process of developing guidelines (NICE, 2005). NICE has defined a 'reference case' in which submission to the institute should follow to ensure consistency of health technology assessment (NICE, 2008c) (NICE, 2013). The reference case suggested by NICE recommends the use of current practice as the best comparator for economic evaluation. Also the cost of treatment or procedures should include all costs pertaining to the NHS or personal social services and all health effects should be

considered. The outcome is measured in terms of the quality adjusted life year (QALYs) which enables comparison across studies and is normally required by NICE. The source of valuing QALYs should be that of the general population as opposed to a patient population since the NHS is acting on behalf of the general population in distributing resources. A discount rate of 3.5% should be applied to costs and consequences (NICE, 2008c) (NICE, 2013).

In Scotland, the Scottish Medicine Consortium (SMC) and Scottish Intercollegiate Guidelines Network (SIGN) both have a national and international reputation for the quality of their output, and provide guidance that is central to the provision of high quality, evidence based health care in Scotland (NHS Scotland, 2013). Unlike SIGN, SMC carries out an evaluation of the clinical and cost effectiveness of all newly licensed drugs that represent good value for money to NHS Scotland and offers advice to health boards and their area drug and therapeutic committees. As the NHS has limited resources, the SMC work to confirm that those drugs which represent good value for money are accepted for routine use so that they can benefit patients. Moreover, SMC also analyses information supplied by the drug manufacturer on the health benefits of drugs.

2.2.1 Rationale for economic evaluation

The increasing cost of health care relative to other sectors, due to the labour intensive nature of health care provision, has led to pressure on health care budgets. Moreover, health care resources are limited by total funds available, as well as through competition with other areas, such as housing, education and finance. This raises the question of how to decide to allocate the money properly (Hine, 1999). Is this health care scheme worth implementing compared with other ways to use the same resources? Are we satisfied that health care resources are spent in one way rather than some other way? (Drummond, 1987). Economic evaluation sets out to answer these questions.

2.2.2 Principles of economic evaluation

Economic evaluation has become necessary a part of modern health care evaluation. New technologies in health care contribute to rising expenditure in health care programs (Alcocer and Cueto, 2008) (Drummond, 1987). Dummond defines economic evaluation as “the comparative analysis of alternative courses of action in terms of both their costs and consequences”. Therefore, the basic principle of any economic evaluation is not just dealing with costs but also in the way that the consequences of health care programs are measured and valued. In fact, both costs (inputs) and consequences (outputs) of

economic analysis may be employed to distinguish and label several evaluation situations commonly encountered in the health care evaluation literature (Drummond, 1987). The distinguishing of characteristics of health care evaluations are discussed based on Drummond and colleagues work (Drummond, 2005b). In not comparing alternatives (where only a single service or programmes is being evaluated), there can be only the consequences of the programs that are evaluated; it can be labeled as an outcome description. On the other hand, cost description deals only with costs, but not full economic evaluations because alternatives are not compared (Drummond, 2005b). If both costs and consequence of a single program are examined, the evaluation is called a cost-outcome & description.

In the comparison of alternative programs, if only the consequences of the alternatives are compared, then the study is called an efficacy or effectiveness evaluation. Cost analysis examines the costs of the alternative programs. However, none of above-mentioned entirely fulfils the requirements of an economic evaluation. Drummond and colleagues named and designated those studies as partial evaluations. These analyses may not represent full economic evaluations but remain important nonetheless as they provide an understanding of either costs or consequences even if they do not provide a full economic evaluations. The four techniques used in a full economic evaluation include the following; cost-minimisation, cost-benefit analysis, cost-effectiveness analysis and cost-utility analysis. They involve the identification, measurement, and followed by the comparison of the costs (input) and consequences (outcome) of two or more alternative programs. Different types of economic evaluations are discussed in Section 2.10.

2.2.3 A framework for conducting economic evaluation

The most well-known and popular classification framework in health economics is given by Drummond and colleagues (Drummond, 1987). They introduced an eight step framework for conducting economics, as shown in Table 2.1. This framework is admirable for its simplicity and transparency. This section is discussed only step 1 to step 6.

Table 2.1 Drummond's framework for economic evaluations

Step	Framework
Step 1	Define the health intervention & Study perspective
Step 2	Identify & Describe the Alternative
Step 3	Identify, Measure & Value all Relevant costs
Step 4	Identify, Measure & Value all Relevant Benefits
Step 5	Discount Future Costs and Benefits
Step 6	Perform a sensitivity analysis
Step 7	Perform a Marginal analysis
Step 8	Make recommendations

Source: Adapted from Drummond et al (2007)

2.3 Study Perspective

A well-designed health economic evaluation should clearly state the study perspective and the interventions under study (Shih and Halpern, 2008). The perspective of an economic evaluation is the viewpoint from which the study is conducted. There are a number of alternative perspectives, including societal, third-party payer, patients, employers, or health care providers. These can be classified into two groups. First, a health sector perspective concerns the healthcare-related cost and impact on the government. Second, a societal sector perspective focuses all impacts inside and outside the health sector. These impacts all have related costs in terms of resource usage, which should be taken into account (Drummond and McGuire, 2007). According to the US Panel on Cost-effectiveness in Health and Medicine, economic evaluations in social perspective should include all the cost and consequence of health care interventions, which concern the health-related costs and impact on the government and on the private sector(such as third-party payers, clients and their families or caregiver) (Russell et al., 1996). Although the societal perspective is recommended in many text books of economic evaluation, the choice of perspective depends on the aim of the evaluation.

2.4 Comparators

Economic evaluation is always structured to compare alternatives. Choosing suitable comparators for analyses is important, as their selection is crucial in order to determine the cost effectiveness of the intervention and the study's relevance. Generally, it can be found that the alternative expected to be replaced if an intervention is adopted will be the comparator.

In the reference case, the most widely used comparator is “current practice”, also referred to as “base case” and “standard of care”. However another common comparator is, “do nothing”, outlined in the Generalized Cost-Effectiveness Analysis (GCEA) framework (WHO, 2003). According to the GCEA guidelines, the appropriate comparator for optimal intervention is a hypothetical “back-calculation” of “what if current practice is eliminated”. This is namely the “null comparator” (WHO, 2003).

Hay and Jackson, however, have stated that a comparator should not just be the current standard of care as a single standard may not exist (Hay and Jackson, 1999). Moreover, the choice of option should include evidence from local guidelines and treatment patterns and may perhaps be based on the review of literature and expert opinion. In other words, in the analysis, a comparison should be made first of all with standard therapy. If standard therapy is not appropriate as a comparator, the most frequent therapy or the most effective therapy can also be chosen. The comparator should not be constrained by the immediate concern of the decision maker, data available or restrictions of current practices. There should be justification provided for the exclusion of any possible options. If good practice is to anticipate future comparators, particularly lower cost intervention, it may lead to an underestimation of the incremental cost-effectiveness ratio (ICER) of the new intervention.

2.5 Costs

Drummond and colleagues state that the cost of heath intervention or treatment can be categorised into four main groups (Drummond, 2005b).

- *C1 are costs arising from the use of resources within the health sector such as medical visits, pharmaceuticals, hospitalisation etc;*
- *C2 are costs forming the use of resources by patients and their families such as travel costs, co-payments for medical services or drugs, expenditure on home modifications specifically related to the illness and its treatment, time cost and other out-of-pocket expenses;*
- *C3 are costs related to resource use in other sector such as welfare organizations, forensic service, educational services etc; and*
- *C4 are productivity losses*

Furthermore, costs can be divided into direct and indirect costs (Shih and Halpern, 2008). Direct costs focus on the resource consumed, related to medical care cost interventions. On the other hand, indirect costs consider the time consumed or

saved by patients and their families who look after the patients as a result of treatment and intervention. Sometimes, indirect costs measure the long-term labour market consequence of disease or interventions (Shih and Halpern, 2008).

Costing in health economic evaluations are dependent on the perspective of a health economic evaluation and the different types of costs are further outlined in Table 2.2. The type of costs to be included in an analysis depends on the aim and perspective of study design. Health economic analyses performed in the health care providers perspective or public programme tend to focus on direct costs only, C1, which might include inventory carrying cost, pharmacy time to compound or dispense, nursing time to administer and even allocated hospital overhead costs. Other perspectives, including third-party payers, businesses, clients and their families or caregivers might be considered in C2 + C3. Analysis done for an employer might consider only costs that affect the employees, for example an intervention that results in productivity loss of workers is classed as C4. In health economics a more complete model, "societal" perspective, which captures the full economic impact of new treatments and interventions is sometimes used, and requires both direct and indirect cost; C1 + C2 + C3 + C4. However, while preferable as an economic theory point of view, this perspective is more complex and time-consuming than the health sector perspective.

Table 2.2 The relation between perspectives of economic evaluations and cost

Perspective				Costs types	Example
Health care provider	Public payers		Societal perspective	C1	Medical visits Pharmaceuticals and medical devices Hospital services Medical services Community base service such as home care Health care providers and other staff Equipment, space , building and associated overhead costs
				C2	Travel costs Co-payments for medical services or drugs Expenditure on home modifications specifically related to the Illness and its treatment Time cost and other out-of-pocket expenses
				C3	Welfare organizations Forensic service Educational services
		Employer		C4	Productivity losses Short-term and long-term absences from work Cost to employer to hire and train replacement worker

Based on page 19, table 2 in Guidelines for the economic evaluation of health technologies: Canada. (CADTH, 2006)

2.6 Outcome measurement

The choice of outcome measure is dependent on the study question and the type of economic evaluation that is being undertaken (Lorgelly et al., 2010). In valuing health outcomes, NICE advises that all health effects should be accounted for within an economic evaluation. If other non-utility based measures capture health effects that are not captured within the utility measure, the effectiveness of treatment on these measures can also be presented (Lorgelly et al., 2010).

2.6.1 Types of outcome measurement

Gold and colleagues state that the most important factors a researcher should consider while choosing the appropriate outcomes are the beneficial outcomes and adverse effects affecting individuals, for instance patients, families, health care personnel or decision makers and society (Gold et al., 1996). This section describes different types of outcome measures that have been used in health economic evaluation. Table 2.3 categorises health outcomes into three different types including clinical outcomes, quality of life outcomes and monetary outcomes.

Table 2.3 Types and outcome measurement in economic evaluations

Types	Outcome measures
Clinical outcome	
-Morbidity (intermediate outcome)	Health endpoints, other clinical measures, prevalence or event, generally expressed in natural units such as number of cancers detected and episode free days
-Mortality (final outcome single dimension)	Estimate life years gained, survival rate and life year saved
Quality of life outcome (Final outcome with multiple dimensions)	
- Disease specific instrument	EORTC QLC-C30 (cancer) ,Arthritis impact measurement scale
- Generic health questionnaire based	Short Form 36 (SF-36), Short Form 12 (AF-12), Sinkness Impact Profile
- Generic health preference based	The EuroQol 5D (EQ-5D), Short Form (SF-6D) and Health Utility index (HUI)
Monetary	As measured in a contingent valuation excise to elicit an individual's willingness-to-pay for an intervention

2.6.2 Clinical outcomes

Morbidity: Intermediate outcome

The first type of effectiveness unit in cost-effectiveness analysis is the intermediate outcomes, such as the number of heart attacks prevented, the number of ulcer prevented and number of cases detected. Some health endpoints can be classified as binary outcomes, which occur or do not occur. An advantage point of using endpoint prevention as the effectiveness measure is that it is closely related to clinical objective. On the other hand, a disadvantage of using the number of endpoints prevented is that, with an intertemporal model(Johannesson et al., 1996).

Mortality: Final outcome single dimension

A saved life is one type of event that is sometimes used as an effectiveness measure. Survival analysis is a statistical method used to analyse data that can be used to estimate the probability distribution of the expected time to an event, which is often mortality but can be a nonfatal event as well (Drummond and McGuire, 2007). The analysis compares all the different activities that reduce the risk of mortality by using lives saved as an effectiveness measure. In the medical area, one, three, five and ten year's survival is often used in this respect. Similarly, using survival may result in similar problems to the morbidity endpoint measure. For example, if five year survival is used, patients may live for 8 or 10 years after treatment, and is an obvious weak point of such an outcome measure. However, at least in the medical area it has been common to use the number of life-years gained rather than lives saved as the effectiveness measure for programmes that reduce the mortality risk. The advantage of this measure is that it takes into account the number of life-years at risk when mortality risk is reduced. Additionally, life-years gained only focuses on survival and does not incorporate quality of life. Comparisons of health programmes that affect both the quality and quantity of life are considered. Effectiveness measures need to incorporate changes in both the quality and quantity of life (Johannesson et al., 1996).

2.6.3 Quality of life: Final outcome with multiple dimensions

Since the World Health Organisation (WHO) defines health as being not only the absence of disease and infirmity but also the presence of physical, mental and social well-being, in health care research and economic evaluation, quality of life has become ever more important (WHO, 1952) . "Quality of life" and more specifically health related quality of life (HRQoL) refer to the physical, psychological, and social domains of health,

which reflects the definition defined by the WHO (Marcia and Donald, 1996). There are three main approaches to describe and measure HRQol which include disease-specific instruments, generic health questionnaire based instruments and generic health preference based instruments.

Disease-specific scales

Disease-specific scales mainly focus on reference values for individual diseases with specific population and specific medical conditions such as the European organisation for research and treatment of Cancer (EORT) Quality of Life Core 30 (QLC-C30), the Rotterdam Symptom Checklist (RSCL), the Functional living index cancer (FLIC), and the Functional Assessment of Cancer Therapy scale (FACT). For instance, EORT QLC-C30 is frequently used to assess HRQoL in various groups of cancer patients (Aaronson et al., 1993). The strength of disease specific scales is that it is good at assessing specific diseases and is typically better accepted by clinicians and patients. On the other hand, it cannot provide a comprehensive measure of health and quality of life, and it is not possible to make comparisons between conditions.

Generic health questionnaires based instruments

Generic health questionnaire based instruments were developed to attempt to capture the HRQol. The concept of HRQoL takes into account patient well-being as expressed by both the physical and psychological (or mental) domains of health. Generic questionnaires include SF-36, SF-12, and Sickness Impact Profiles. For example SF-36 or its reduced version SF-12 is the most widely used instrument for assessing HRQol in individuals with chronic disease. The SF-36 is a generic multi dimensional instrument for capturing quality of life with 8 dimensions that can be broken down into two broader classes applicable to different populations and comparisons can be made across conditions or programmes; physical and mental (Ware et al., 1993). The eight aspects of health status include: general mental health, physical and social functioning, physical and emotional role, bodily pain, vitality, and general health perception. Scores on each scale range from 0 (worst) to 100 (best) (Brazier et al., 2002). Generic measures can be used on a variety of disease areas or health problems and can be used to compare HRQoL across diseases (Valderrabano et al., 2001). This instrument may not however pick up slight changes in HRQoL which are of importance to the patient.

Generic preference based instruments

Several generic preference based measures have been developed and commonly use questionnaires including the EuroQol 5D (EQ-5D), Short Form 6D (SF-6D) and the Health Utilities Index (HUI)(Whitehead and Ali, 2010). Those values are translated into a zero to one scale where zero denotes death status and one with perfect health. For instance, EQ-5D is a popular method usually used in CUA studies (Richardson and Manca, 2004). The EQ-5D questionnaire has five dimensions including mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These dimensions include three levels which individuals are asked to respond to; whether they have no problems, some problems or severe problems (King, 1996). These preference based instruments also differ in terms of the valuation method: time trade-off (TTO) is used to value EQ-5D, whereas SF-6D and HUI use the standard gamble (SG), and the visual analogues scale (VAS) is also used in HUI (Whitehead and Ali, 2010). Each technique is discussed in greater detail in the next section.

2.6.4 Measuring preferences

The three most widely used techniques to measure the preferences of individuals for health outcomes, and so derive QALY weights are the rating scale, also called the visual analogues scale (VAS), the time trade-off (TTO) and the standard gamble (SG). Those methods are described below.

Visual analogues scale (VAS)

VAS is the simplest approach in which people are asked to rate health states on a scale (ranged between 0 and 100). The 0 represents the worst possible state, assumed to be equivalent to death, and 100 is the best possible state, or perfect health. The SF-6D questionnaires are used to derive utilities. The questionnaires are comparatively easy to administer and the results are easy to generalise across disease areas (Bravo Vergel and Sculpher, 2008). However, when compared with preferences measured by TTO and SG, this technique may not act as an interval scale of preferences (Bleichrodt and Johannesson, 1997).

The time trade-off (TTO)

The second method of measuring quality weights is TTO. It was developed by Torrance and colleagues (Torrance et al., 1972). In the TTO approach, respondents are asked to choose how many years of full health they would trade off against living 10 years with their current health state. They are given two choices between living the rest of their life (t) in a given health state (i) or a shorter period of time (x) living in perfect health. The

time (x) is varied until the subject is indifferent between the two alternatives, at which point the required preference score for state (i) is (x/t) (Bravo Vergel and Sculpher, 2008).

The standard gamble (SG).

In the SG technique, participants are asked to choose between the certainties of an intermediated health state (such as current health state with a chronic disease) and the uncertainty of treatment. Treatment offers two possible outcomes, one is more attractive than the certain outcome (e.g. curable outcome) and another outcome is less attractive (e.g. death). The quality weight of a health state is assessed by comparing a specific number of years in the health state to a gamble with a probability (p) of achieving full health for the same number of years and a complementary probability (1-p) of immediate death. The probability of full health (p) is varied until the individual is indifferent between the alternative, and the quality weight of the assessed health state is equal to (p) (Johannesson et al., 1996).

Drummomd and colleague stated that future health outcomes are clearly uncertain in the real world. Therefore, the preferences measured under uncertainty (utilities) are the more appropriate (Drummond and McGuire, 2007). The three different methods result in different scores. The SG generally produces higher utility values than the TTO (Bravo Vergel and Sculpher, 2008). The SG usually gives the highest weight followed by TTO and VAS.

2.6.5 Valuing life years: the concept of a QALY

The outcomes from treatment and other health programs have two fundamental components, the quality and the quantity of life (Whitehead and Ali, 2010). The quality of life will result in a number of dimensions relating to both physical and mental capacity. The quantity of life is expressed in terms of survival or life expectancy (Phillips and Thompson, 2009). A Quality adjusted life year (QALY) takes into account both the quantity and quality of life generated by health care interventions and treatments. It is the arithmetic product of life expectancy and measure of the quality of the remaining life-years. Moreover, a QALY combines both mortality and morbidity measures of health by weighing a year of life by the quality of life (that is utility) experience. QALYs are estimated by assigning every life-year a weight between 0 and 1. A weight of zero reflects a health status that is valued to be equal to death and one reflects full health. There are also disability adjusted life years (DALYs) as an outcome measure (Lorgelly et al., 2010). DALYs are years of healthy life lost, 1 represents the greatest amount of

disability (equivalent to death) and 0 represents no disability (full health) (Gold et al., 2002).

The concept of QALYs is to combine the survival of an individual with their HRQoL, as shown in figure 2.1. This demonstrates the QALYs that can be gained by an individual by comparing one treatment and without treatment. The QALY shows a concept of how many extra months of years of life of a reasonable quality a person might gain as a result of treatment or interventions (particularly important when considering treatments for chronic disease). Consider an individual suffering from a particular disease without treatment, the lower line shows the health profile if no treatment is received, where the HRQol reduces over a period of time until death (Death A). In treatment groups, their HRQol remains at a higher level for longer. However, the subject would experience side effects during treatment and then be without disease until they die (Death B).

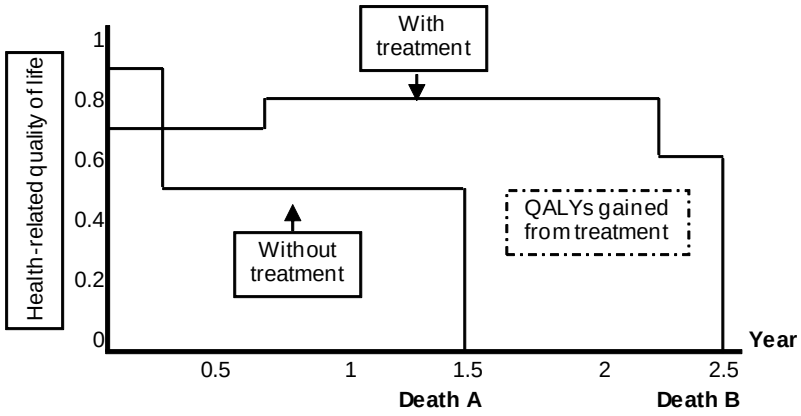


Figure 2.1 QALY gained from an intervention
Based on fig 6.6 chapter 6 in Methods for the Economic Evaluation of Health care Programmes (Drummond, 1987)

For example, The QALY is calculated simply by multiply the duration of time spent in a health state by the HRQoL weight (utility score). The QALYs calculation based on this figure is as follows. Without treatment the estimated QALYs are 0.25 year at 0.9 + 1.5 years at 0.5 = 0.98 QALYs. With treatment the estimated QALYs are 0.75 year at 0.7 + 1.5 year at 0.8 + 0.25 year at 0.6 = 1.88 QALYs. Therefore, 0.9 QALYs are gained with treatment.

Traditionally, the impact of health care has been measured in terms of its effects on mortality. Since, health is much more than just being alive, its effects on morbidity are increasingly being taken into account. Moreover, mortality and morbidity data says nothing about what the priority weights given to one condition relative to another ought

to be. Therefore, to quantify the health effects of therapy, it is necessary to measure something about their effects on life expectancy and health related quality of life. CEA expresses the outcomes (effectiveness, benefits or consequences) in terms of natural units, which include survival, cost per exacerbation avoided, improvement in health status, years of life gained or clinical parameter (e.g. response or remission rates, reduction in blood pressure etc). The results are usually expressed as a cost per unit of effect. However, in any one program there is often more than one outcome of interest, some outcomes are more important or more valued than others. Intermediate outputs are unsuitable because they cannot directly be converted into an outcome measurement like QALYs gained, which is required for CUA.

2.7 Time Horizon

The time horizon should be sufficient in length to include all major clinical and economic outcomes and is distinct from the length of treatment effect (Gold et al., 1996) (Husereau et al., 2013). However extending the time horizon beyond the period after which there are no meaningful differences is unnecessary and to maintain consistency, costs and outcomes should have the same time horizon applied to them.

The time horizon essentially depends on the clinical area being considered, however lifetime horizons are suitable for most models, and for decision tree models this would normally be from the onset of treatment until a recovery or death (Hay and Jackson, 1999). Shorter time horizons can be considered for acute conditions, for example; however this should only be done if there are no long term complications associated with the disease process or intervention . Finally, in certain circumstances multiple time horizons may be justified, for example trivial data can be collected for short term analysis in order to complement more important patient data collected over the long term (NICE, 2013).

2.8 Discount Rate

Economic evaluation combines total costs and health outcomes over time. However if such costs and outcomes occur beyond one year in the future they should be discounted to present values to adjust for time preference. A standard method for discounting should be adopted across evaluations, and the discount rate clearly stated, as discount rates are not universal across all guidelines, with differences depending on setting, location and perspective of an analysis. For example, NICE set the discount rate at 3.5% for both costs and outcomes (NICE, 2008c) (NICE, 2013).

2.9 Sensitivity analysis

Sensitivity analysis is an essential tool to investigate outcomes that are obtained from health economic models. The tool is employed to estimate the effect of uncertain variables on the robustness of model results, and indeed Adronis (2009) argues that sensitivity analysis is an appropriate method for handling uncertainty (Andronis et al., 2009)

There are several ways of undertaking sensitivity analysis, such as one–way and multi-way sensitivity analysis and probabilistic sensitivity analysis, each of which is discussed below (Briggs et al., 1994). Together these complementary tools offer a collection of techniques for dealing with uncertainty, and are best used together to help clarify the level of uncertainty in the results of a model, and consequently of the model robustness and its recommendations.

2.9.1 Univariate (one-way) sensitivity analysis

This is the simplest form of sensitivity analysis and involves simply varying one value in the model by a given amount, and examining the impact that the change has on the results in the model. One-way analysis can be performed using a number of different approaches, each of which is useful for different purposes (Briggs et al., 1994). For example, each parameter of a model can be changed individually by a certain amount to find out which parameters have the greatest effect on the results of the model. Additionally another approach would be to individually vary every parameter in the model to their greatest and least possible values, based on the data confidence intervals, in situations where the decision maker has little confidence in the models inputs. The limitations of one-way sensitivity analyses should be noted. The examination of one parameter at a time creates a incomplete picture of uncertainty and may underestimate how uncertain the results of the model actually are. Finally, a more detailed approach to univariate sensitivity analysis, called ‘threshold analysis’, is to examine the impact of varying the size of a single input parameter over a range of values. This helps the analyst to understand the relationship between the model’s results and the input values (Briggs et al., 1994).

2.9.2Multivariate (multi-way) sensitivity analysis

This is an extension of one-way sensitivity analysis, which involves changing two or more key parameters at the same time and studying the combined effect on the results of the evaluation (Briggs et al., 1994). This type of analysis takes into consideration the

probability that more than one parameter value in a model is uncertain, however it becomes increasingly more complex and difficult as the number of inputs involved increase (Taylor, 2009). One advantage of multivariate sensitivity analysis over one-way sensitivity analysis is that its results are more realistic; one-way analysis can be difficult to interpret as it only presents a partial picture of uncertainty.

A specific form of multivariate sensitivity analysis is 'scenario analysis', of which there a number of approaches. For example extreme case analysis, also known as 'worst/best' case analysis looks at a set of extreme values across parameters; the most optimistic and pessimistic values chosen from the perspective of the intervention being assesse (Briggs et al., 1994)d. If an intervention is preferred under base-case assumptions, high cost/ low effectiveness (pessimistic) assumption, low cost high effectiveness (optimistic) assumptions, the analyst can be confident in the conclusion of the study (Briggs and Gray, 1999).

2.9.3 Probabilistic sensitivity analysis

The final form of sensitivity analysis that is considered is probabilistic sensitivity analysis (PSA). This method assigns a specific distribution to each input parameter in the model rather than a single value and, and by drawing randomly from those distributions using computer software, a large number of mean cost and effectiveness estimates are generated that can then be used to form a joint distribution of the differences in cost and effectiveness between interventions (Briggs et al., 2006). This sensitivity analysis method can provide a more complete assessment of the uncertainty associated with all inputs in a model. However, Although PSA is similar to univariate and multivariate sensitivity analysis as parameter values are still substituted with different values, it should complement not replace these sensitivity analyses as PSA does not allow for the consideration of each type of uncertainty (Briggs et al., 1994).

2.10 Type of Economic Evaluation

The aim of economic evaluation is to present a methodology, which can be employed to certify that the efficient deployment of scarce health care resources in the absence of the price signals a free market would produce. Economic evaluation is essentially a comparative technique that examines alternative courses of action in terms of both their costs and health outcome consequences (Drummond, 1987). The selection of a type of economic evaluation depends on the purpose of the study or the question to be addressed and may also influence by factors such as data or target audience.

There are several different types of economic appraisal. There are different ways in which these types can be classified and presented. The first method is to classify them according to efficiency. The second is to classify them according to cost and benefits measured and how this is done. Lastly, economic appraisal can be classified by the type of decision that they apply to. The most well-known and popular classifications in health economics are given by Drummond and colleague (Drummond, 1987) (Drummond, 2005b). Drummond et al classified using both the measurement and decision type principles which include cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA). Similarly, four different types of economic evaluation are often discussed in the literature; cost-minimization analysis (CMA), cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) (Cunningham, 2001) (Shih and Halpern, 2008). Furthermore, the possible fifth approach is cost-consequence analysis (CCA) has been employed by other commentators (Drummond, 1987).

This thesis presents the types of economic evaluation in two broad classifications, distinguished by whether or not health outcomes are measured in the same units as resource consequences. Hence CBA describes the situation where all coast and benefit are measured in the same units (usually monetary). CEA is used to explain the general class of approach where costs are measured in terms of money, but that health outcome is measured in units other than money. CUA is considered as a special case of CEA characterised by a generic measure of health outcome. Subsequently, this section briefly discusses types of economic evaluations and the criteria used to select the types.

2.10.1 Cost-minimisation analysis

Cost-minimisation analysis (CMA) is a method of economic evaluation in which two or more therapeutic alternatives are compared in terms of net cost, which chooses the least total cost alternative interventions. CMA is applied when the evidence is shown to have equivalent clinical outcomes of two alternative programmes. Moreover, the transparent and intelligible equivalence of comparators in terms of efficacy must be presented.

2.10.2 Cost-benefit analysis

The second analytical type is cost-benefit analysis (CBA). CBA is an evaluation technique that quantifies both costs and outcomes, compares alternative interventions using in monetary units such as the Dollar or the Euro. Subtracting the costs of an

intervention from the value of outcomes of the intervention can calculate the net benefit. The choice criteria as to whether or not to accept the intervention or technology is considered whether the benefits are greater than the costs subject to budget limit.

The main advantage of CBA as an analytic technique is that is has an uncomplicated decision rule which is consistent with economic theory and will lead to an efficient distribution of health care resources. Since resources and health outcomes are valued in the same units, if the health outcome benefits are greater than the resource use cost then the health care programme in question should be implemented. If, on the other hand, the costs outweigh the benefits, the programme should not be implemented. The drawback of the CBA is that a monetary assessment of clinical results must be made even though methodologically this is difficult to present. Because of those methodological difficulties, this method of analysis is not used for NICE decision making (NICE, 2013).

2.10.3 Cost-effectiveness analysis

The main purpose of CEA is to maximise the health effects obtained for a given budget (Karlsson and Johannesson, 1996).The definition of cost-effectiveness analysis (CEA) is taken to be the analysis of alternative courses of action in terms of their resource costs, evaluated in monetary terms and health outcome effects evaluated in natural units(NICE, 2008c). The more useful of CEA is a full economic evaluation where both costs and outcomes of health programmes or treatments are analysed. CEA expresses the consequences (effectiveness, benefits or outcomes) in terms of natural units, which include survival, cost per exacerbation avoided, improvement in health status, year of life gained or clinical parameter (e.g. response or remission rates, reduction in blood pressure etc) (Oostenbrink et al., 2005) (Shih and Halpern, 2008). CBA is appropriated for answering questions about whether or not health programmes or treatment are an efficient use of resources, while CEA presents which of the possible ways of providing those is the most efficient. However, CEA is a handy technique for comparing alternative intervention whose effects are in similar units but can not be used to measure an isolated single programme and cannot evaluate interventions which have several types of clinical effects. In this problem are drawbacks that lead to development of cost-utility analysis (CUA).

2.10.4 Cost-utility analysis

The final type of economic evaluation is cost-utility analysis (CUA) where costs are measured in monetary units and the benefit is assessed as a not-monetary but utility-

adjusted outcome, the quality adjusted life year (QALY). On the other word, the benefit measurement is an extension of CEA that identifies the change in health status measure known as the QALY. Information from QALY along with cost can be applied to guide resource allocation. Such an approach allows decision to be made concerning the suitable level of resources to be devoted to different health care programmes. Hence, CUA analysis can be seen as addressing allocative efficiency concerns within the health service budget. The implicit goal of CUA is the maximisation of health gain from a fixed budget. CEA and CUA are similar methods of economic evaluation, particularly in the USA and some authors do not distinguish (Gold et al., 1996).

2.11 Using decision analytic modelling for economics evaluation

Over the last two decades, decision modelling has been widely used in many countries around the world to evaluate alternative health care interventions (Hunking et al., 2001) (Taylor et al., 2004). In the health care field, modelling is usually applied to scheduling budgets, the workforce and the location of facilities (Brennan and Akehurst, 2000). In addition, decision analysis is an established framework used to inform decision making under conditions of uncertainty. Policy making involves choosing an action after weighing the costs, benefits and risk factors of the information of individual patients or the patient population (Siebert, 2003). The purpose of decision modelling is mainly to allow for the variability and uncertainty associated with all decisions (Briggs et al., 2006). Subsequently, the results of decision analysis can inform a decision both for an individual and for health care policy (Eddy, 1990) (Richardson and Detsky, 1995). Models for the economic evaluation of health technologies make valuable information available to policy makers and are a useful tool for demonstrating the detailed and complex “real world” with a more simple and understandable structure (Taylor, 2009), represented by a series of numbers, mathematics and statistical relationships. Several definitions have been presented for the term “model” as it applies to the context of health care. The International Society for Pharmacoeconomics and Out-comes Research (ISPOR) Task Force on Good Research Practice defines Modelling Studies as “an analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources, and whose purpose is to estimate the effects of an intervention on values of health consequence and cost” (Weinstein et al., 2003). However, it is unlikely that a model will ever include all the possible ramifications of a particular option under consideration (Briggs et al., 2006).

Decision analysis is applied flexibly to the evaluation of health care, where alternatives are judged upon the basis of expected costs and health outcomes. Decision analysis is a systematic approach that uses explicit and quantitative methods to analyze decisions under conditions of uncertainty (Pauker and Kassirer, 1975). It uses mathematical relationships to define a series of possible consequences and result in a set of alternative options being evaluated. Based on the inputs, the likelihood of each consequence is expressed in terms of probabilities where each consequence has an outcome and a cost. The technique can calculate the expected cost and outcome of each option under evaluation. The purpose modeling in economic evaluations is to synthesise data for the purpose of making a decision.

2.12 A framework incorporating decision analytical modelling into economic evaluations

There are many stages in developing a decision model for economic evaluation (Briggs et al., 2006). In 1976, Shepard and Thompson suggested measuring costs in monetary units and health benefits in non-monetary units, consisting of five major analytical steps (Shepard and Thompson, 1979). However, the most prevalent framework that incorporates decision analytical modelling for economic evaluations was proposed by Biggs and colleagues (Briggs et al., 2006). The six distinctive stages of this framework, presented in Table 2.4, indicate how to develop a decision model when conducting an economic evaluation. The first step begins with defining the programme or intervention, which focus on processes and limits. Secondly, both net monetary cost for the prevention and treatment of the illness under the proposed programme or intervention, based on the cost of current practice, as well as the discounted present value are computed. The third step is to compute the respective health effects or benefits, followed by a sensitivity analysis, which is performed in steps four and five. The final step is then to complete a Value of Information analysis. This framework does not however quantify the decision uncertainty associated with the mean health outcome and mean cost in the model.

Table 2.4 Briggs's framework for decision analytical modelling

Step	Framework
Step 1	Specifying the decision problem
Step 2	Structure the decision model
Step 3	Identify and Synthesis evidence
Step 4	Deal with Uncertainty and Heterogeneity
Step 5	Presenting Uncertainty in cost, effects and cost-effectiveness
Step 6	Value of additional research

Source: Adapted from Briggs et al (2006)

2.13 Specifying the decision problem

This step entails clearly identifying the question that needs to be addressed in the analysis (CADTH, 2006). The perspective, comparators of study, and time horizon of the model, in terms of which cost and consequence are considered relevant, should be stated clearly (Philips et al., 2006). Moreover, disease states should be selected to reflect the underling biological process of the disease in question and the impact of interventions and therapies (Philips et al., 2006).

2.14 Structure the decision model

Economic evaluation can take different structures including formalised approaches such as decision tree, Markov analysis, and discrete event simulation. Models are useful tools for simulating real-world situation in health economic evaluations. The choice of structures depends on the nature of disease, the impact of technology and the availability of data for its assessment (Brennan et al., 2006) (Brennan and Akehurst, 2000). Briggs and colleagues also suggested that choices have to be made to structure the model based on the nature of the interventions themselves and the natural history of a particular condition and the impact of the options on that process (Briggs et al., 2006). There are two main approaches in the construction of a model for the CEA, namely deterministic and stochastic approaches. Where decision tree and Markov analysis is deterministic approach, and discrete event simulation (DES) and Monte Carlo simulations is stochastic approach. The majority of the health care models use a deterministic approach (Drummond, 2005b) (Drummond et al., 2005). On the other hand, Briggs and colleagues consider the two forms of the cohort model. First, whether the model should seek to characterise the experience of the “average” patient from a population sharing the same characteristics, such as decision tree and Markov models. The other form of the model should explicitly consider the individual patient and allow for variability between patients, which use the micro simulations model (Briggs et al., 2006). This thesis present the decision analysis based on Briggs and colleagues.

2.14.1 Decision trees

The decision tree is probably the most common structure that describes a clinical decision and possible outcomes (Drummond, 1987) (Briggs et al., 2006) . A decision tree is a graphical model that shows the consequence for each possible outcome. A decision tree is included with three types of nodes: decision nodes, chance nodes and terminal nodes, shown as squares, circles and triangles respectively. Typically, the decision tree is drawn by starting at the far left with decision nodes and moving from left

to right through a consequence decision and chance nodes. The decision tree diagram begins with the decision node, which represents the alternative options or interventions of the decision in the model. Chance nodes may come out of the decision nodes, showing the range of possible pathways. Similarly, the pathways are built up through a series of branches showing each pathway. A branch of chance nodes will illustrate the probability of a particular pathway occurring. Probabilities show the likelihood of a particular event at the chance node. Moving from left to right, the probabilities in each branch of those following events relate to those patients who have undergone particular previous occurrences. Each pathway ends in the terminal node which represents the average outcome of this event. The expected value probabilities are calculated by multiplying along the pathway probabilities of subsequent events, which is the joint probability. Expected values are based on the summing up of the pathway values weighted by the pathway probabilities. For economic evaluation, the costs and health benefits are shown in the final outcomes in each partway. Another way of working out the expected costs and effectiveness is by 'rolling back' the tree for given option in a decision tree. This will give exactly the same answer as the approach mentioned above.

However, simple decision tree is usually appropriated with situation which time frame is short and the mortality of patients does not differ across strategies (Barton et al., 2004). The tree is not suitable in chronic disease and can become very bushy in modelling, making them unwieldy and complex to programme and analyse (Drummond, 1987) (Drummond and McGuire, 2007).

2.14.2 Markov Model

Markov models have been frequency applied for published decision analyses (Barton et al., 2004). The limitation of decision trees models are resolved by Markov models but certainly not all. Particularly, Markov models are functional when the decision problem involves a risk that is ongoing over a period time. Markov models can deal with the pattern of recurring –remitting disease over a period of time and of competing clinical risks characteristic of many chronic diseases, which is needed to reflect a large number of possible outcomes over time (Briggs et al., 2006). A cohort-based Markov matrix uses a transition probability per unit time for individuals in the cohort to change to another state, with related costs and utilities. Subjects in cohorts are designed to be in one of a finite number of health states and a patient in each state can only make one state transition during a cycle. The cycle of models have durations, depending on the nature of the disease, which might last one day, one week, one month or one year. Subjects move through the model starting at the beginning, middle or end of each cycle. The patients in the modelling move through the models in a number of ways, either stay in

the same state, or travel to a worse state (or a state in which a specific event occurs), or they move to death state. The transition probabilities determine the proportion moving between states at fixed time periods or cycles (Glasziou and Sanders, 2002). Cost and effects are typically incorporated into these models as a mean value per state per cycle. Adding the cost and effectiveness over the states and then weighting them in accordance with the time the patient will be in each state gives expected values

Markov models are applied to estimate the clinical and economic consequences for disease with recurrent events. The limitation of Markov modelling is that the model lacks memory (Briggs et al., 2006). Using the Markov model on a patient in a certain state, the model cannot make a distinction between someone who has just entered a state and someone who has been in the state for some time and remains in the same state. A problem with this is that time is frequently correlated to worsening disease, especially with chronic disease and this correlation is ignored in this model. In addition, decision trees are used within Markov model as call Markov cycle trees (Vos et al., 2005). One form of this is when transitions between Markov states are characterise in term of a tree (Briggs et al., 2006).

2.14.3 Discrete Event Simulation

Discrete Event Simulation (DES) is unlike Markov Modelling in several aspects. DES can resolve the problem associated with the Markov model where there is no memory of previous events. The three components included in DES are entities, event and time. Entities usually incorporate personal characteristics into modelling. Individuals' attributes such as age, sex, previous history of illness of subject can be taken into account and tracked over the modelling period through the natural history of disease. An event is anything that occurs during the simulation, such as an exacerbation, and events can occur sequentially, simultaneously or both. This idea extends well beyond the transitions in Markov model because the event need not imply a change in the patient's state. In terms of time, it is an explicit element of DES models and the model can be run for as long as is necessary are more flexible than Markov models (Le Lay et al., 2006). The DES is not restricted to the use of equal cycle length, but when using Markov modelling a fixed cycle length is required. Also, the time to the next event may derive from either a parametric or an empirical distribution to characterize the variability of time to next event between individuals in the population (Karnon, 2003). DES is most successfully applied to processes such as modelling the introduction of new equipment within the setting of a GP clinic or for queuing. Furthermore, two main interactions between individual patients should be taken into account in two main circumstances. First, when modelling of infectious disease, where the risk of an individual catching the disease depends on how

many other people already have it. Second, when restrictions on resources mean that the choice of treatment for one patient affects what can be given to another. Methods such as DES and system dynamics (SD) are required when interaction is a significant issue in modelling. DES works at an individual level, where as SD works on an aggregated level (Barton et al., 2004).

2.15 Identifying and synthesizing evidence

For economic evaluation it is important to use all relevant evidence. In the context of parameter relating to the effectiveness of intervention, this is consistent with evidence-based medicine. Ideally, evidence on outcome should come from a randomised controlled clinical trial. If a clinical trial is unavailable, then evidence should be obtained from a meta-analysis, or an indirect treatment comparison analysis. In the UK, the NICE demand a systematic review, evidence-based analysis of input parameters in the model. This framework is required to synthesise this range of evidence (NICE, 2008c) (NICE, 2013) . Moreover, this also requires synthesis of evidence in a computer-based mathematical model. The role of decision analysis is a methodology for synthesising evidence on different parameters to inform estimates of net cost and net effectiveness from different sources (Karnon and Brown, 1998). Clinicians, health care managers, policy makers and consumers have wide-ranging information needs. They require high-quality information on the effectiveness, meaningfulness, and appropriateness of a large number of healthcare interventions. Moreover, economic evaluation also needs appropriate information for using as an input parameter in an economic model. Data can be chosen based on the principle of systematic review and meta-analysis

2.16 Making decision models probabilistic

Economic evaluation models, as mentioned in section 2.12, are a useful tool to aid policy makers to make decisions in health care programmes. Economic evaluation of interventions or therapy often relies on mathematical models to account for relevant clinical, biological, epidemiological, and economic factors. The key point of decision modelling is to permit for the variability and uncertainty linked with all decisions (Briggs et al., 2006).

2.16.1 The role of probabilistic models

The aim of probabilistic modelling is to be a sign of the uncertainty in the input parameters of the decision model and explain what this means for uncertainty over the outputs of interest such as measures of cost, effect and cost-effectiveness, in both

incremental cost-effectiveness ratios and net-benefit measure (Briggs et al., 2006) (Claxton et al., 2002). There will never be complete information on all the possible consequences of a particular intervention in a given population. This uncertainty needs to exist in order for decisions to be made according to the best available evidence and best practices in decision analytic modelling (Bilcke et al., 2011) . Briggs and co-workers propose three main reasons why it is important to consider uncertainty, even if the concern of decision maker is expected values. Firstly, most models relate to combining input parameters in ways that are additive and also multiplicative as power functions, resulting in models that have non-linear input and output parameters. Secondly, uncertainty over the results of an analysis implies the possibility of incorrect decision making which impose a cost in terms of benefits forgone, such that there may be value of obtaining more information even in a world where our only interest is in expected values. Lastly, policy changes are rarely costless exercises and decision reversal may be problematic, such that there may exist value associated with delaying a decision that may be impossible, or problematic to reverse (Palmer and Smith, 2000) (Briggs et al., 2006). This section will describe how models can be made probabilistic in order to capture parameter uncertainty.

2.16.2 Variability, heterogeneity and uncertainty

Uncertainty exists in all economic evaluation. There are three concepts commonly referred to in relation to uncertainty in economic evaluation; variability, heterogeneity and decision uncertainty.

Variability refers to the differences that are found amongst patients by chance, and in some of the medical literature, is referred to as first order uncertainty. On the other hand heterogeneity refers to the differences amongst patients that can mostly be explained, such as age and sex. As heterogeneity is not a source of uncertainty though, it is quite distinct from variability.

Additionally, rather than variability or heterogeneity, it is uncertainty that decision models aim to capture. A number of guidelines for economic evaluation discuss the importance of considering different types of decision uncertainty and specific frameworks for quantifying uncertainty have emerged (Briggs and Gray, 1999) (Briggs, 2000). Two forms of uncertainty can be identified, the first of which is parameter uncertainty. Economic models require information to populate them, and these information requirements are often referred to as the parameters of the model (Briggs, 2000). The level of uncertainty surrounding a parameter is dependent on the extent to which its value has been estimated and has been referred to as second order

uncertainty to differentiate it from variability (Briggs et al., 2006). The second form of uncertainty is model uncertainty, is considered to be external to the decision model and does not relate to the parameters but to the assumptions formulated by modelling framework (Drummond, 2005b) (Kim et al., 2010) (Bilcke et al., 2011). In other words, it refers to uncertainty about the extent to which the structure of the model adequately captures the relevant characteristics of the disease and intervention being investigated (Gold et al., 1996).

Stinnet and Paltiel refer to the categorisation of uncertainty in terms of first and second order (Stinnett and Paltiel, 1997). They explain that first order uncertainty (or variability) is random error, which reflects the inherent stochastic natural of a trial. This uncertainty persists even when the probability distribution of an outcome is known with certainty. Second order uncertainty is systematic error that corresponds to uncertainty in the parameters of the probability distribution of the outcome. In CEA, second order uncertainty is often considered. There is usually a lot of uncertainty in the estimation of expected health costs and outcomes, which leads to uncertainty around estimates of intervention respective, ICER.

2.16.3 Distributions for probability parameters

In probability sensitivity analysis, inputs are defined as probability distributions to reflect their full uncertainty. In capturing parameter uncertainty in the estimation of expected values of a parameter, the model needs to represent the sampling distribution of the mean. This has important implications for the choice of distribution for any of the parameters to represent the uncertainty in any parameter of the model, which vary depending on the family or from of distribution (Briggs et al., 2006). However, the probability distribution that represents uncertainty in a decision analytic model is not chosen randomly but decided based on the type of data, the parameter type and estimation process (Claxton et al., 2005b) (Briggs et al., 2006). Briggs and Claxton and colleagues discuss the types of distribution as follows:

Normal distribution: The normal distribution is a very important statistical data distribution pattern. When graphed as a histogram it creates a bell shaped curve known as a normal curve. The curve is symmetrical about a single central peak at the mean (μ) (average of the data), so fifty percent of the distribution lies to the left of the mean and fifty percent lies to the right. Therefore, a random variable (parameter) on the normal distribution is of any value between negative and positive infinity ($-\infty, \infty$). Because of the central limit theorem, such distributions are considered to be a good choice to represent uncertainty. The standard deviation (σ) describes the amount of variation of the normal

distribution. The mean value of a parameter and its standard deviation are required to calculate normal distribution.

Log normal distribution: The log normal distribution is a continuous probability distribution that ranges from zero to positive infinity $(0,\infty)$and is positively skewed. This distribution does not cover any negative values and this distribution is suitable for parameters that are non-negative as well as highly skewed or are multiplicative, such as ratios. A normal distribution is generated if the natural logarithms of the parameters of such a distrirbution are calculated.

Beta distribution: In probability theory and statistics, the beta distribution is employed to model the proportion of successes (n) in a binomial trial and define the interval (0,1). Two positive shape parameters that appear as exponents of the random variable and control the shape of the distribution are α and β. α is the number of events that occur and β is the number of non-events.

Dirichlet distribution: The Dirichlet distribution, being multivariate in nature, is generally considered to be the multinomial extension of the beta distribution with one parameter per category. It is thought of as flexible and convenient, computationally, as its components take values (0,1)

Gramma distribution: The gamma distribution is useful for continuous variables, particularly those considered to be highly skewed. It is constrained within the interval zero to positive infinity $(0, \infty)$. Gamma distribution is represented by α and β.

2.16.4 Choosing distributions for parameters

This section discusses how to choose distributions for parameters of decision models. Assigning the probability distribution is the first step in making a probabilistic model. The common distributions in PSA are Normal, Log-normal, Beta and Gramma distributions. Table 2.5 shows the common types of distributions and distributional forms of parameters (Briggs et al., 2006).

Probability parameters can only take values between zero and one (0,1). Beta distributions are commonly representative of such parameters as prevalence, diagnostic test accuracy, and the probabilities exclusive events must some to one. Uncertainty in this probability can be presented by two parameters, α and β, as mention in the previous section, β = (sample size (n) – the number of events occurring (α)). Normally, α and n are reported in publications, and these are used to calculate the β for the beta

distribution. On the other hand, as opposed to binomial data, multinomial data with numerous categories, each represented by proportions that sum to one, are appropriate for the Dirichlet distribution. If the overall sample size and the number of events of interest for each category are reported in a publication, the data can then be used to calculate proportions for each category to fit the Dirichlet distribution.

Costs data are calculated from resource usage, weighted by unit costs. Therefore, cost parameters should not be lower than zero, as it is not possible to have a negative result, although they can range up to infinity. When considering which distribution best fits the costs parameter, gamma distribution can be selected as it is constrained to value zero and upward to infinity, $(0, \infty)$. If cost estimates of a suitably large sample are found to be symmetric around the mean, it can be assumed that the central limit theorem applies and a normal distribution fitted to the data. However if the data is highly skewed, a more characteristic situation for cost data, both log-normal and gamma distribution are used to fit the distribution.

Utility can be suited to the Beta distribution only if it is appropriate to assume that the utility range is between close to zero and close to one. However, in cases of severe life-threatening illness, the utility can be very low or negative utility and so utility ranges between negative infinity and one $(-\infty, 1)$. Therefore, the beta distribution should be avoided in such cases. When the transformation of UD = 1 - utility, where UD is a utility decrement, the distribution is constrained on the scale zero to infinity $(0, \infty)$ and is better fitted to a Gamma or Log-normal distribution.

Where publications only report the mean and standard deviation (SD) for a parameter point estimate, the Method of Moments can be applied to obtain α and β from the known values of the mean and SD in order to fit the Beta distribution. In situations where publications report the mean without standard error (SE), an assumed standard error that is of sufficient size to reflect a broad enough uncertainty range can be adopted. However if rather than a specified SE, a 95% confidence interval is reported, then the SE can be calculated using the confidence limits.

Table 2.5 Distribution parameters and the distributional forms

Distribution	Scale	Cost	Utility	Relative risk	Prevalence	Mortality	Treatment probability
Normal (continous)	(-∞,∞)						
Lognormal (interval)	(0,∞)						
Gramma (continous)	(0,∞)						
Beta (binomial)	(0,1)						
Dirichlet (multinominal)	(0,1)						

2.16.5 Assessing parameter uncertainty

Parameters represent the point estimate (mean values) for a particular population or subgroup, but they are estimated with uncertainty from sample data. In order to handle this parameter uncertainty in the decision model, there needs to be an assessment of how it impacts the result on the analysis. Cost-effectiveness analysis based on the mean value is called deterministic analysis. In probabilistic models, the uncertainties around individual parameters are represented using an appropriate distribution for each parameter and by drawing a random estimate of the distribution to represent the point estimate. By propagating this uncertainty though the model, a measurement of the uncertainty in the outcome statistic is derived. A parameter simulation technique (Monte Carlo simulation) is used to assess the implication of the results of the study of uncertainty in all of the inputs. Using simulation techniques allows propagating parameter uncertainty through the model.

Once the simulation process has been undertaken for propagating the uncertainty, the situation is comparable to one where sample data is available and provides an estimation of mean overall parameters such as cost and effectiveness of therapy. The average of cost, effect and cost-effectiveness are used to present the probability outcomes of uncertainty in the input parameters by drawing randomly from the parameters numerous times. In the best practical guidelines (e.g. the BMJ, NICE and US Panel), it is recommended that the uncertainty surrounding estimates of cost-effectiveness needs to be explored when presenting economic evaluation results (NICE, 2008c). Using any number of iterations greater than 1,000 times is acceptable (Briggs et al., 2006). The next section will discuss how to present the cost effective results.

2.17 Presenting the results

This section reviews in detail the cost-effectiveness plane (CE plane), followed by a discussion of the decision rules for cost-effectiveness analysis. Once the relevant parameters in a decision model are determined and the appropriate distributions applied, probabilistic sensitivity analysis can be performed. Random values form each parameter distribution are sampled randomly using Monte Carlo simulations with 1,000 iterations as mentioned in the previous section in order to provide different cost, effect and cost-effectiveness outcomes for each vector of the input parameters. NICE guidelines for methods of technology appraisal 2013 suggest standard deciding factors should be followed when combining costs and QALY, these should reflect when

dominance or extended dominance exists, and present a thorough incremental cost-effectiveness ratio (ICERs) (NICE, 2013).

Cost-effectiveness analysis on the CE plane

The purpose of cost-effectiveness analysis is to compare the cost and health outcomes of one treatment compared to some relevant alternatives (O'Brien et al., 1994a) (Shepard and Thompson, 1979) (Sculpher et al., 1997). In other words, a new experimental therapy (or treatment group) may be compared with and some current practice (or control). In terms of costs, the true costs of the new therapy (C_t) versus the control therapy (C_c) are presented and the true effectiveness of the new therapy (E_t) versus the control therapy (E_c) are also illustrated, as shown in Table 2.6. Four situations have been identified by O'Brien and colleagues that can occur with regard to the incremental cost and effectiveness of therapies, as outlined in (O'Brien et al., 1994b) (Drummond and McGuire, 2007).

Table 2.6 Characteristics of cost-effectiveness and not-cost-effectiveness regions in the cost-effectiveness plane (CE)

Situations	Interpretation	CE plane quadrants
1. $C_t - C_c > 0$; $E_t - E_c < 0$	dominance – reject experimental therapy as it is both more expensive and less effective than existing therapy	NW
2. $C_t - C_c > 0$; $E_t - E_c > 0$	trade-off consider magnitude of the additional cost of the new therapy relative to its additional cost	NE
3. $C_t - C_c < 0$; $E_t - E_c < 0$	Trade-off consider magnitude of the cost-saving of the new therapy relative to its reduced effectiveness	SW
4. $C_t - C_c < 0$; $E_t - E_c > 0$	dominance – accept experimental therapy as it is both cheaper and more effective than existing therapy	SE

Base on page 174, chapter 8 of economic evaluation in health care (Drummond and McGuire, 2007)

Four situations in are equivalent to the four quadrants of the CE plane. This plane has been advocated for the analysis of cost-effectiveness results (Anderson et al., 1986). The difference in cost and the different in health outcome between two therapies can fall into one of four quadrants of the CE plane. There are a lot of authors discussed and presented the CE plane in different way. The four situations are presented in different ways. The most popular used is compass direction which shown in north, south, east and west (Hoch et al., 2002) (Fenwick et al., 2006) (Briggs, 2007). Another

way to present the result in different quadrants of the cost-effectiveness plane is Roman numbers (Black, 1990) (Briggs, 1998) (Drummond, 2005b).

This thesis presents the cost-effectiveness plane by compass direction. Figure 2.2 illustrates the cost-effectiveness plane. In the diagram, the horizontal axis represents the difference in effect between the therapies of interest, and the vertical axis represents the difference in costs. A therapy can be placed anywhere on this diagram according to its incremental costs and effectiveness. In the SE and NW quadrants one intervention is simultaneously cheaper and more effective than the other (situation 1&4). However, quadrants NE and SW (situation 2&3) on the CE plane represent where an intervention is both more effective and more costly. A trade-off must then be made between the additional health outcomes and the additional resources required. Or, in other words, a judgement needs to be made concerning whether the additional cost of the more expensive therapy is justified by the additional effectiveness associated with that therapy.

In order to summarise this trade-off, an incremental effectiveness ratio (ICER) is calculated. If this ICER is less than the maximum acceptable cost-effectiveness ratio then the treatment is considered cost-effective. The ICER is discussed below. A straight line is drawn across the NE and SW quadrants and through the origin (as shown in Figure 2.2) and represents the maximum acceptable cost-effectiveness ratio. The line divides the CE plane into two parts; cost-effective and non-cost-effective. The right side of the line indicates that therapies are cost-effective, while the other side indicates cost-ineffective therapies.

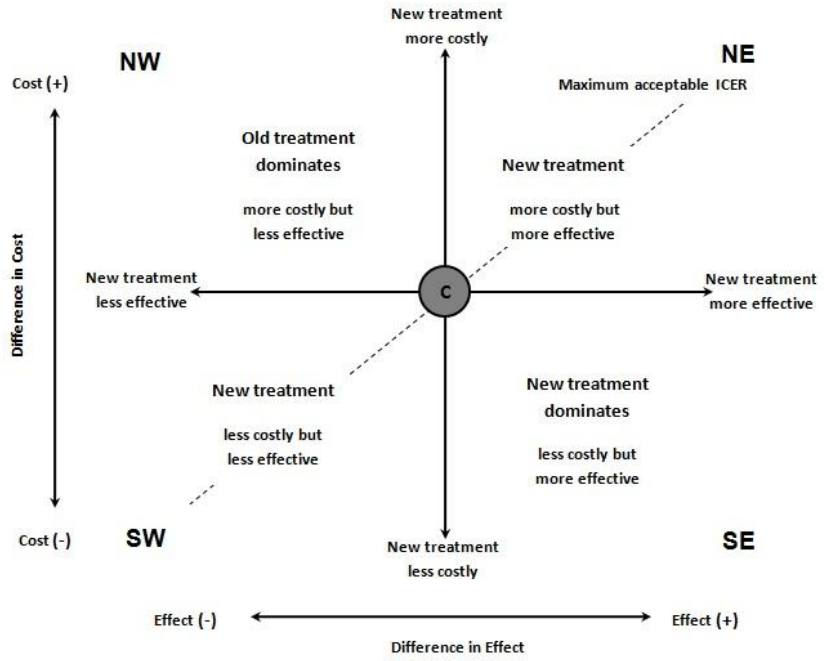


Figure 2.2 The cost-effectiveness plane
Based upon figure 2.3, Uncertainty in cost-effectiveness of health care intervention (Briggs, 1998)

Incremental cost effectiveness ratio

The results of CEA are summarised in a cost-effective ratio (CER) (Phillips, 2009). According to the individual programme, the result may be an average cost-effectiveness ratio (ACER). An independent programme requiring ACER is calculated for each programme, by comparing total costs and total outcomes. The ACER can be calculated by equation below. However, ACER does not provide guidance in decision making and is inappropriate to maximise the health effects for a specific share of resources (Karlsson and Johannesson, 1996). Therefore, ACER is inconsistent with the underlying decision rules of cost-effectiveness analysis.

$$\text{ACER} = \frac{Ct}{Et} \text{ and } \frac{Cc}{Ec}$$

It is likely that choices will normally have to be made between different therapy options for the same condition, different dosages/treatment compared with prophylaxis. What must be questioned however is the amount of benefit that can be achieved from a new therapy and what cost it is associated with? To answer this question incremental cost-effectiveness ratios (ICER) are used. ICER can be calculated, which obtains a summary of the cost-effectiveness of one intervention compared to the other. Similarly, ICER is the difference in costs between the two interventions divided by the difference in

their consequences, and can be described as the incremental price of a health unit outcome from the intervention, compared to the other. The interventions that have a low cost-effectiveness ratio are good value and would be the preferred option. Alternatively, where the costs and benefits of each alternative are calculated and compared with their next best alternative, rather than with a common alternative. The ICER can be calculated by the following equation.

$$\text{ICER}=\frac{\text{Ct}}{\text{Et}}-\frac{\text{Cc}}{\text{Ec}}\neq\frac{\text{Ct-Cc}}{\text{Et-Ec}}$$

The ICER determines the appropriate measure of cost-effectiveness rather than the ACER. The difference of two average ratios is not equal to ratio of the differences (Briggs et al., 2006). The first problem with ACER, in order to make the judgement concerning whether therapy represents good value for money, decision-makers need to consider the additional cost of new therapy, as treatment A, over the control therapy, as treatment B, in comparison to the additional effect, that is the ICER. Suppose that we are comparing treatment A versus treatment B. If Treatment B gives the lowest cost per unit effect, treatment A shows the greatest overall effect. It could be that society wishes to provide treatment A rather than treatment B even though its average cost per unit effect is higher. Another problem with ACERs is that they give no information on the relative position of the two treatments in the cost-effectiveness space.

Incremental net benefit

In this section, two methods for developing point estimates for the difference in cost and effect, the cost-effectiveness ratio and net monetary benefit (NMB) are described. If both cost and effectiveness are higher with the new intervention, the question will come up with will decision maker, should a new intervention be accepted as cost effective or not. To answer this question, the decision rule is applied. Decision rule requires the decision maker to know the maximum amount that the payer would be willing to spend for and additional benefit or maximum acceptable cost-effectiveness ratio. In the other words, the maximum willingness to pay for a benefit is known as the cost-effectiveness ceiling ratio, which can be plotted as a line through the origin on the CE plane, as demonstrated in figure 2.3. The ICER remains the most popular method of presenting the result of CEA and CUA (Drummond, 1987). Interpretation of ICER requires the choice of a cost-effectiveness ceiling where representing the maximum that society would be willing to pay for an incremental health benefit, and the development of the decision rule based on this maximum. If the estimated ICER is below some maximum willingness to pay for

health gain then that intervention represents good value for money. For example, if the ceiling ratio was £ 30,000 per QALY gained, then an intervention which presents incremental costs of £10,000 and increased QALY by 0.4. Thereby, it would have an ICER at £ 25,000 per QALY and would be considered to be cost-effective in comparison to the alternative. If the ceiling ratio presents at £20,000 per QALY, then an ICER of £25,000 per QALY would not be considered cost-effective. The cost-effectiveness decision rule rearrangement by ICER can be calculated by equation 2.3. The maximum acceptable willingness to pay or ceiling ratio is denoted by λ.

$$\text{Decision-rule}=\frac{\text{Ct-Cc}}{\text{Et-Ec}}<\lambda$$

One of the most major drawbacks of the ICER relates to the mathematical difficulty in creating confident intervals for a ratio (McFarlane and Bayoumi, 2004). The net benefit approach employs a simple re-arrangement of the cost-effectiveness decision rule of equation 3 in order to overcome the problems with cost-effectiveness ratios. The limitation of ICER have led some authors to prefer the incremental net monetary benefit (INMB) (Hoch et al., 2002) (McFarlane and Bayoumi, 2004). The INMB is calculated by the increase in effectiveness multiplied by the amount of the maximum acceptable willingness to pay and subtracting from this the incremental cost of achieving the benefit, as shown in equation below. Using the NMB approach, a therapy is deemed to be cost-effective if a positive incremental net-benefit suggests that the therapy represents good value for money, while a negative side suggests the intervention is cost-ineffective (Drummond, 2005b). >

$$\text{INMB}=\left(\text{Ct-Cc}\right)\lambda-\left(\text{Et-Ec}\right)>0$$

The INMB expression is probably most familiar to economists, and is the one most often referred to as “net benefit” in the literature, as mention above. Thus, one potential advantage of INMB compared with cost-effectiveness ratios is that we can directly compare the difference in arithmetic mean monetary benefits between treatment groups to determine NMB by use of the same types of univariate and multivariable methods. Moreover, using the net benefit approach, it is also possible to re-arrange the inequality in another way to define the net health benefit (NHB), as shown in equation below.

$$NHB=\left(Et-Ec\right)-\frac{(Et-Ec)}{\lambda}>0$$

For both expressions, a positive incremental net benefit presents the health gain greater than that from investing the same resources in an alternative therapy. A negative incremental net benefit suggests the intervention is cost-ineffective.

2.18 Decision making, uncertainty and the value of information

The previous section described how models can be made probabilistic in order to capture parameter uncertainty. In recent years there has been considerable emphasis on the development of an appropriate method for handling uncertainty in mathematical models, with a trend to move from univariate sensitivity analysis towards a fuller probabilistic description of uncertainty, e.g. cost-effectiveness planes, cost-effectiveness acceptability curves (CEAC) and distribution of incremental net benefits (Tappenden et al., 2004). This section will briefly address how the results of probabilistic modelling should be interpreted by decision makers, and how to address the question of whether more evidence is required.

2.18.1 Decision making with uncertainty

Once distributions have been applied to each of the appropriated parameters in decision model, the probabilistic sensitivity analysis can be performed. The probabilistic analysis explores uncertainty in cost-effectiveness outcomes by using Monte Carlo simulation (1000 iterations) in order to determine expected costs, outcomes and cost-effectiveness. The minimum of 1,000 iterations of Monte Carlo simulation is used to sample random values from each parameter distribution simultaneously to provide different cost, outcome and cost-effectiveness for each vector of input parameters. Those outcomes across all 1,000 iterations represent the probability outcomes. Incremental cost and incremental effectiveness for each of the 1,000 iterations can be represented visually using a cost-effectiveness plane (CE plane). The joint distribution of the costs and effectiveness from the Monte Carlo simulations (1000 iterations) are plotted on a cost-effectiveness plane in order to show the impact of uncertainty in the model parameters on the model outcomes in both expected incremental costs and effects (Briggs, 2000). 95% confident intervals (uncertainty intervals) can be calculated by using the lower percentiles (0.025) and upper (0.975) percentiles from the simulations results using the

percentile method (Briggs et al., 2006). Uncertainty in the incremental outcomes is demonstrated when the results cross over the y-axis, representing both QALY gains (in the eastern quadrants) and QALY losses (in the western quadrants), as shown in Figure 2.3. Similarly, a spread through the origin passing through the x-axis represents uncertainty in the incremental costs of the intervention. The size of the spread also shows the extent of uncertainty in the costs.

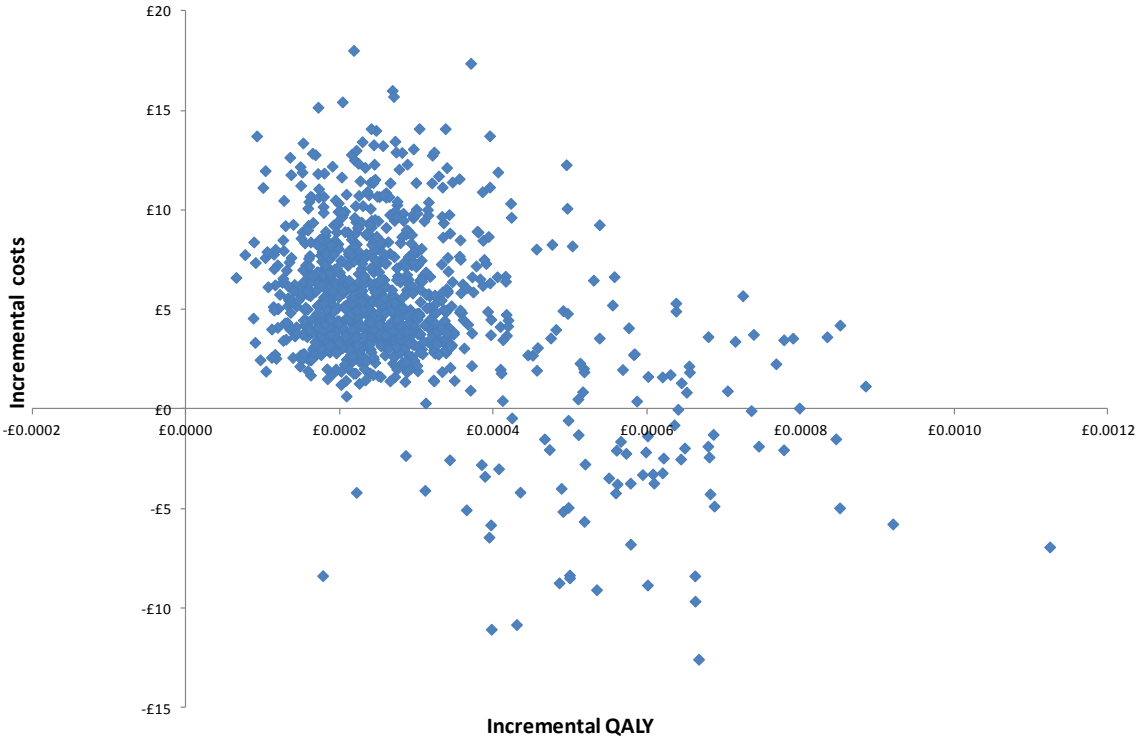


Figure 2.3 The CE plane

The uncertainty surrounding the cost-effectiveness of a technology, for a range of thresholds for cost-effectiveness, can be resulted as a cost-effectiveness acceptability curve (CEAC), which illustrate the uncertainty surrounding the estimate of cost-effectiveness and were developed as a result of considerable debate regarding the best way to deal with such uncertainty (Briggs and Fenn, 1998) (Fenwick and Byford, 2005). The information from a CEAC should not be used to make statements about the implementation of the intervention or therapy (Fenwick and Byford, 2005). The uncertainty around the cost-effectiveness estimate is illustrated in Figure 2.4.

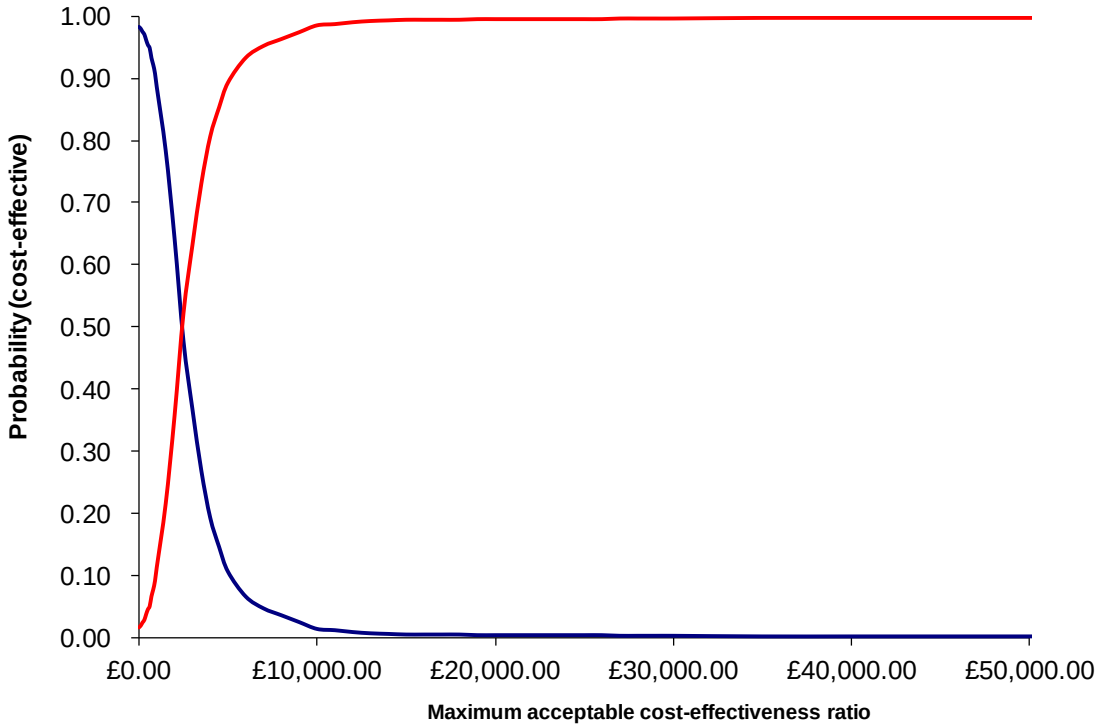


Figure 2.4 The CEAC

Figure 2.4 illustrates a hypothetical CEAC to demonstrate the probability that each intervention is cost-effective at different thresholds. At a threshold of £10,000 there is a 95% probability that the new treatment is cost-effective and an only 5% probability that old treatment is cost-effective. The new intervention presents as the greatest expected net benefit, therefore it would be considered the optimal choice at this ceiling ratio. At a threshold at £5,000, new and old interventions are cost-effective with 50%. There remains a 50% probability that new and old treatments are the wrong choice, which is the uncertainty in the decision.

The CEAC is straightforward to calculate and interpret and is therefore an ideal technique for presenting uncertainty in the cost-effectiveness outcome from a probabilistic sensitivity analysis to decision makers who have to make the choice of whether to adopt or reject a new intervention, based on the current evidence. Uncertainty over the results of analysis leads to the likelihood of incorrect decision making. In order to adequately address that the new treatment is cost-effective that old treatment, it is necessary to consider two issues, namely, knowing the decision uncertainty and current evidence, should the technology be adopted or not and is additional research essential in order to support the decision (Briggs et al., 2006). Presenting the probability sensitivity analysis with CEAC can be considered to be the

first question (the technology should be adopted or not), in addressing the second question, further techniques are required.

2.18.2 Value of information

Having constructed a decision analytic model, subsequently undertaken probabilistic sensitivity analysis and then considered decision uncertainty, the results of the probabilistic sensitivity analysis can, in addition, be used to perform a value of information analysis (VOI).

Bayesian decision theory and the value of information method provides an analytic structure which can address both whether a technology should be adopted based on current evidence and whether more evidence is required to support this decision in the future (Claxton et al., 2002) (Briggs et al., 2006). Likewise, value of information (VOI) analysis has been suggested as a systematic decision analytic approach for aiding decision makers in assessing whether there is enough evidence to support and adopt new therapies, and setting research priorities for health technology assessment (HTA) (Eckermann et al., 2010). VOI is based on the reasoning that decisions based on information that already exists will be uncertain, with this uncertainty there is therefore a chance that a wrong decision is made, presenting a cost in terms of implications for the health of patients who receive less than optimal care and less than efficient use of health care resources.

Decisions based on existing information will be uncertain, and there will constantly be an opportunity that the mistaken decision will be made. Although, decision makers make the right choice based on our current estimate of expected net benefit, there is a chance that another therapy would have had high net benefit once our current uncertainties are resolved. For example, at threshold of £10,000 there is a 95% probability of being cost-effective in new intervention as shown in Figure 2.4. New intervention would be considered the optimal choice as it has the greatest expected net benefit [ENBt >ENBc]. While control intervention is the optimal choice, 5% of the time it would have been the wrong decision and this represent uncertainty in the decision to adopt the new treatment. The probabilistic sensitivity analysis results have been presented in previous sections and the decision uncertainties for each of the analyses have also been explored, however to demonstrate the probability that each intervention is cost-effective at different thresholds it is necessary to consider two issues, namely, knowing the decision uncertainty and current evidence, should the intervention be adopted or not and is additional research essential in order to support the decision

(Briggs et al., 2006). If such questions are left unanswered, it may be the case that decision makers attempt to interpret the cost effectiveness results in terms of how to make the decision to adopt or reject the intervention given the uncertainty. Uncertainty about the results of an analysis can lead to incorrect decisions, and this can have a cost in terms of the correct decision that was not acted upon. (Claxton, 1999). Decision makers wish to evade incorrect decisions, and so there is a value in obtaining more information, if it can reduce uncertainty.

In the UK, the application of the value of information approach was demonstrated for health technology assessment (HTA) programme and for the UK reimbursement decision body by NICE, however, practical application of EVPI in publish literature are limited (Claxton et al., 2004) (Claxton et al., 2005a).

EVPI per decision/patient

In the ideal world in which there is perfect information, the most favourable intervention (and the most cost-effective) would be the chosen each time. On the other hand, in a non ideal world where there is uncertainty, on choosing an intervention ordinarily thought to be optimal there is a strong possibility that it will be a less favorable one. The expected opportunity loss associated with this uncertainty can be interpreted as the expected value of perfect information (EVPI), as the perfect information can get rid of the probability of making the incorrect decision or it is the amount the decision maker should be willing to pay to eliminate all uncertainty in the decision. With estimates of the probability of error and the opportunity cost of error, the expected cost of uncertainty or the expected opportunity loss surrounding the decision can be calculated (Briggs et al., 2006).

The EVPI was calculated using the probability of cost-effectiveness for each strategy, and were generated in the CEAC calculation within a range of values at intervals of £500 from zero to £100,000/QALY. As discussed in the previous section, net benefit can be determined for each of the interventions for every one of the thousand iterations of the Monte Carlo simulation (Treatment (NBt) and Control (NBc)). The average is then taken in order to calculate the expected net benefit for each iteration, so that the optimal strategy can be determined. With existing evidence, the optimal intervention is that which has the greatest expected net benefit for all 1000 Monte Carlo iterations [max(ENBt : ENBc)]. The expected net benefit for the optimal strategy is the value that is assigned to the decision made based on current information. Given perfect information, the optimal strategy would be the decision of choice each time, and by choosing the strategy that maximizes net benefit for each iteration of the Monte Carlo

simulation, this can be accounted for: from iteration one through to iteration 1000: [max(NBt1 :NBc1],[max(NBt2 : NBc2)] and ….[max(NBt1000 : NBc1000)]. The Mean of the 1000 optimal net benefit choices is the expected value of the decision based on perfect information [E[max(NBt : NBc)]]. The expected value of perfect information is therefore the difference between the value of a decision with perfect information and without perfect information.

$$\mathrm{EVPI} = \mathrm{E}\left[\max\left(\mathrm{NBt}:\mathrm{NBc}\right)\right] - \max\left(\mathrm{ENBt}:\mathrm{ENBc}\right)$$

EVPI population level

Having estimated EVPI per patient, the population EVPI for one year is calculated using population estimates. In order to determine the population value for EVPI, the patient population over the lifetime of the technology must be taken into account in terms of the relevant patient population who would benefit from the screening tests. Representing the EVPI per decision in terms of the patient population, the population value for EVPI, gives an idea of the upper limit for expenditure on future research into the decision question. It is calculated by multiplying the patient population (both present and future) that is able to benefit from the information, by the difference between a decisions expected value (the greatest net benefit expected) with current information and with perfect information (Briggs et al., 2006).

The process of calculating the EVPI follows on from calculating the EVPI per decision. Firstly, the annual patient incidence (I) for the specific disease is used to represent the annual patient population. The expected timeframe of the intervention can then be estimated in years (t) and a discount rate applied. For example, the annual pregnancy population is estimated to be 60,000 mothers (NHS, 2013). With regard to screening tests for GDM, this study assumes an effective technology life of 10 years with new patients eligible for treatment each year. An appropriate timeframe for the intervention should be used for which estimates of the the effectiveness of the technology used in the model are unlikely to change and are relevant. Given an annual disease incidence in population (I) and the intervention lifetime of T years (t), then the effective population can be calculated by applying a discount rate (r) for patients in future years, and summing the population across the years (t). Consequently, the effective population is multiplied by the EVPI per decision to give the population level EVPI (EVPIpop). The EVPI pop equation is detailed below.

$$\text{EVPIpop} = \text{EVPI} * \sum_{t=1,2,\ldots t} \frac{\text{It}}{(1+r)^t}$$

The EVPI can be used to point out whether future research is likely to be worthwhile. However, it does not take into account the cost of future research. This may be construed as a drawback to EVPI, however it is still useful in providing an indication of whether future research is worthwhile or not and is an important part of VOI analysis.

EVPI for parameters (EVPPI)

Where the EVPI analysis indicates further research is worthwhile, the next step involves identifying what type of research should be performed. In order to reduce uncertainty in the cost effectiveness decision, it is necessary to consider what parameters are driving uncertainties in the model. In this respect, the expected value of perfect parameter information (EVPPI) is used to distinguish parameters for which it would be valuable to have more accurate estimates. Briggs and colleague use EVPPI as an abbreviation for the expected value of perfect information for a parameter (Briggs et al., 2006). Similarly, Ades and colleague denote EVPPI as expected value of partial perfect information (Ades et al., 2004). Those addressed in the same meaning of the difference between expected value with perfect and current information about the parameter.

The EVPPI is calculated from the difference between the maximum net benefit calculated from current estimates that involve some uncertainty and the expected net benefit that is calculated from partial perfect information. This provides the proportion of general uncertainty furnished by any single or group of parameters and gives a guide to where research should be focused for greatest efficiency. The equation of EVPPI is detailed below. The EVPPI algorithm with the parameters of the model denote by (θ) including the perfect parameter of interest and the other parameters which keep their distribution from the probabilistic sensitivity analysis. The net benefit of an intervention (t) if the parameters of the model take the value (θ) is denoted by NB (t, θ).

$$\text{EVPPI} = E_\theta \left[\max t \left(E_\theta NB(t,\theta) \right) \right] - \max \left[E_\theta NB(t,\theta) \right]$$

The various steps in the EVPPI process are detailed as follows:

1. Firstly, a parameter of interest is chosen for which perfect information is required, then a random value is taken from its probabilistic distribution and held constant, to represent ‘perfect’ information for the parameter of interest.

2. The Monte Carlo simulation is run once more maintaining the ‘perfect’ parameter as a constant but allowing the probabilistic draws from all other parameters to run.

3. Then the average NB under treatment and control (average NBt) (average NBc) from the 1000 iteration Monte Carlo simulation is subsequently recorded in addition to the intervention identity that gives the maximum expected net benefit [max(NBt :NBc)].

4. Following the Monte Carlo simulation a second random draw is made for the ‘perfect’ parameter of interest, so that a new ‘perfect’ value can be held constant. Step 1 through to 3 are repeated 1000 items, each time holding a different value for the perfect parameter estimate constant while the other parameters in the Monte Carlo simulation are allowed to vary. For each Monte Carlo simulation the mean net benefit for treatment and control is recorded along with the intervention identity that gives the maximum net benefit.

5. Once the process has been completed, the 1000 stored mean NBs for treatment and control and maximum intervention identities are used to calculated the expected net benefit (ENB) for each intervention [ENBt , ENBc] and the expected maximum net benefit [E[max(NBt:NBc)]] across the 1000 Monte Carlo simulations outcomes

6. The intervention with the greatest ENB [max(ENBt :ENBc)] is the expected value of a decision based on current information; the intervention which has the greatest ENB and would therefore be the optimal (cost-effective) choice.

7. The expected maximum benefit [E[max(NBt:NBc)]] is the average of the 1000 maximum net benefit interventions from each of the Monte Carlo simulations. This is the expected value of perfect parameter information.

8. The final stage in the EVPI process is to subtract the ENB of the decision under current information from the ENB of the decision with perfect parameter information which gives the expected value of perfect parameter information (Briggs et al., 2006).

In terms of which parameters would add the most value in this additional information and how it should be collected, the expected value of perfect parameter information (EVPPI) is used; understanding what drives the uncertainty is necessary to establish the type

and perhaps the scale of future research. The EVPPI analysis reports results in terms of the value per patient and these are presented in terms of the pertinent patient population that may benefit from this supplementary information. Parameter uncertainty for one or a group of parameters can be eliminated instead of the EVPI for all parameters being estimated at once.

EVSI

Alternatively, expected value of sample information (EVSI) is related with predicting the expected reduction in uncertainty resulting from the collection of data from an additional finite sample (Tappenden et al., 2004). In other words, EVSI represents the reduction in uncertainty that may be expected to result from further information from studies with predetermined sample size. As mentioned previously, the expected value of perfect information (EVPI) places an upper limit on the returns to further research. This deals with a necessary condition for carrying out future research, where more research about the problem as a whole, or about particular parameters or parameters groups, may be worthwhile if the EVPI or EVPPI is more than the cost of conducting more research. Nonetheless, to establish an adequate research design, we need to consider the marginal benefits and marginal costs of sample information (Briggs et al., 2006). EVSI, the net of the costs of sampling provides the expected net benefit of sampling or the societal pay off to proposed research.

2.19 Conclusion

This chapter addresses the introduction to the subject of Health Economic Evaluations with a focus on details of Economic Evaluations. Key concepts of economic evaluations have been given in terms of the cost, study perspective, time horizon and discounting rate. In the economic evaluation it is crucial to select the appropriate comparators of the analysis. The time horizon should be long enough to capture all important difference in cost and consequence in both the comparator and interventions. In the cost process, all relevant costs should be considered. From a UK perspective, this should represent all costs to the NHS and social services. The measuring of utility is a vital component of economic evaluation and the QALYs are proposed as the measure of HRQoL for valuing intervention and treatment. The QALYs is preferred because of simplicity, clarity, ease of application, and face validity. Cost and consequence (QALYs) in two different interventions can be compared in order to create a cost utility value, ICER. There is different structure modelling in economic evaluation including formalised approaches such as the decision tree, Markov analysis, and discrete event systems. The majority of

cost-effectiveness studies are based on cohort modelling. Cohort modelling provides a flexible framework which can be programmed relatively easily and evaluated rapidly. The decision tree and Markov model are the two basic types of decision model that predominate in health economic evaluation. Decision analysis allows different strategies to be compared in terms of expected outcome. All relevant events and possible complications are considered along with their probabilities and relevant clinical outcomes and costs are compared. In a situation where information is plentifully, decision analysis can be used to synthesise the best available data from various sources, in order to present results which are as unbiased and realistic as possible.

Cost-effectiveness analysis has become wildly used in evaluation technique, which relates costs and health outcome effects in different units in terms of a ratio. It is important to select a structure that is appropriate for the disease under study and the question or purpose of the economic evaluation study. The CE plane is a useful device for clarifying the concept of cost-effectiveness. The CEAC indicates the probability that the intervention is cost-effective compared with the alternative, given the data and for a given value of the maximum acceptable ratio (λ). Uncertainty should be considered in every stage of economic evaluation, if at all possible, using probabilistic sensitivity analysis. It is not just parameter uncertainty that is important, however, the importance of model uncertainty should be considered at the same time. Use of cost-effectiveness acceptability curves is a method for summarising information on uncertainty in cost-effectiveness. Furthermore, the analysis is on estimating expected value of parameters, the normal distribution is always a candidate distribution because of the central limit theorem. The choice of distribution should be informed by the logical constraints on the parameter and the form of the data/estimation process. Expected value of information (EVI) analysis should be incorporated into the decision problem because it provides an important framework for determining the expected payoff of conducting further research to resolve the model uncertainties.

Chapter 3.The diagnostic test evaluation and application of Bayes' theorem to diagnostic test evaluation

3.1 Introduction

The aim of this chapter is to introduce a general overview of diagnostic test evaluations and consider the approach of Bayes' theorem to diagnostic test evaluation. The chapter begins by describing the fundamental terminology of diagnostic test evaluations. Next, the difference between screening and diagnostic tests are discussed, in terms of purpose, target population, test methods, positive results and costs. The following section discusses the clinical performance of laboratory tests, distribution of the tests and cut-off points. After that, sensitivity, specificity, predictive values, and likelihood ratios are also discussed. Diagnostic tests produce test results, which can divide patients into two groups, disease present and disease absent. Likewise, clinicians can measure sensitivity and specificity and how often specific symptoms occur with disease and without disease, but cannot directly measure the predictive value of a set of symptoms (positive predictive value and negative predictive value). Consequently, this chapter gives an introduction to the theorem of Bayesian inference which can address this problem and in relation to this discusses how clinicians can use sensitivity and specificity, the quantities that a clinician can estimate to calculate predictive values, and which clinicians need to make appropriate diagnoses or decisions. The final section will discuss and introduce the combination of test results.

3.2 Diagnostic tests

This section provides a brief description of diagnostic characteristics. First, the principles of diagnostic test evaluations and their basic notation are demonstrated. This is followed by a discussion of test performance and test accuracy and then a demonstration of moving the cut-off point is given. Finally, the trade-off between maximising sensitivity and specificity is briefly discussed.

3.2.1 Characteristics of diagnostic tests

Screening tests are a medical strategy used to detect a disease in individuals in the general population without clinical signs or symptoms of that disease, and who are at sufficient risk of the disease. Diagnostic tests are medical procedures used to confirm positive screening tests and distinguish healthy individuals from people who have the disease (Public Health Action Support Team 2010). Some of the key differences such as purpose, target population and positive result threshold are shown in Table 3.1. The five main purposes of screening and diagnostic tests are as follows: to verify a diagnosis in symptomatic patients; to screen for disease in asymptomatic patients; to provide prognostic information on patients established to have the disease, to monitor the benefits and side effects of therapy; to confirm disease absent patients (Jyoti and Richard, 2009) (Edgar. et al., 1999). In this way, early intervention and management might reduce complications and mortality from a disease. Although screening may result in an earlier diagnosis, not all screening tests have been shown to benefit the person being screened. Over-diagnosis, misdiagnosis, and the creation of a false sense of security are some potential adverse effects of screening and diagnosis (Laking et al., 2006).

Table 3.1 Differences between screening and diagnostic tests

	Screening Tests	Diagnostic tests
Purpose	To detect potential disease indicators	To verify presence/ absence of disease
Target Population	Large numbers of asymptomatic but potentially at risk individuals, or personal history, race/ ethnicity, disease state, or other factors	Symptomatic individuals to establish diagnosis, or asymptomatic individuals with a positive screening test
Test method	Simple, acceptable to patients and staff	Invasive, expensive but justifiable as necessary to verify diagnosis
Positive result threshold	Chosen towards high sensitivity not to miss potential disease	Chosen towards high specificity (true negative) and high sensitivity (true positive). More weight given to accuracy and precision than to patient acceptability
Positive result	Essentially indicates suspicion of disease (often used in combination with other risk factors) that warrants confirmation	Result provides a definite diagnosis
Cost	Cheap, benefit should justify the cost since large numbers of people will need to be screened to identify a small number of potential cases	Higher costs associated with diagnostic test may be justified to establish diagnosis

Based on table 3.3.1 in diagnosis and screening: differences between screening and diagnostic tests, case finding (Ruf M. and Morgan O., 2008).

Test results may help physicians to make a diagnosis in a symptomatic patient (diagnostic testing), or identify disease in an asymptomatic patient (screen testing) (Jyoti and Richard, 2009) The most common tests provide results along a continuous or quantitative scale

(eg. blood glucose level, white blood cell count). The clinicians often use these ranges to diagnose a condition by classifying them as positive or negative and disease present or absent, based on the criteria or cut-off point. (Laking et al., 2006) Diagnostic techniques allow clinicians to allocate the right treatment to the right patient. However, the errors of diagnostic tests or misdiagnosis may risk useless and possibly harmful treatment, or prevent or delay access to beneficial treatment (Laking et al., 2006).

3.2.2 Diagnostic test performance

To show the performance of the screening test and diagnostic test, the fundamental element is the test result which is often used to evaluate the accuracy of the outcome. Screening and diagnostic test results are shown by classifying patients into two groups: one population with disease (or conditional), the other without disease. (Khamis, 1987)

The clinical performance of a laboratory test can be described in terms of diagnostic accuracy. The four outcomes of the test performance are shown, according to the test results, as positive (T+) or negative (T-), and according to whether the disease was truly present (D+) or absent (D-). Table 3.2 displays the status of the person being tested in the columns and the test results in the rows. For example, in cases where there is disease and the test is positive, the outcome is classified as true positive (TP). In cases without disease, but where the test result claims that disease is present, the outcome is classified as false positive (FP). On the other hand, cases without disease where the test confirms its absence are classified as true negative (TN). In cases where the patient has disease, but the test indicates that they don't, the outcomes are classified as false negative (FN). Information on the accuracy of screening and diagnostic tests can be put into a two way table. A two by two table is the easiest and clearest way to calculate and summarize all the information about diagnostic tests.

Table 3.2 Two way classifications of results according to tests and disease status

	Disease					
Test	**Present D+**	**n**	**Absent D-**	**n**	**Total**	**n**
Positive T+	True Positive (TP)	a	False Positive (FP)	b	All positive	a + b
Negative T-	False Negative (FN)	c	True Negative (TN)	d	All negative	c + d
Total	All diseased persons	a + c	All non – diseased	b + d	All tested persons	a + b+ c+ d

3.2.3 Distributions of test results

The test results can be used to allocate people in the population into two groups: those suspected of having the disease, and those thought to be without it. Typically, the test results include quantitative results (e.g., white blood cell count in cases of suspected infection) followed by some type of distribution curve. Test results are random variables and hence are subject to a distribution, which can be, but are not necessarily, a normal distribution. The distribution of test results shows a different mean for patients with or without disease. The variation in results for patients with the disease is quite large, but a high proportion of test results are close to the mean value. A very similar pattern exists for patients without the disease, although the mean value differs.

There is some overlap in test results for the two patient groups, as can be seen in Figure 3.1. Patients with the disease are represented in the distribution to the right, and patients without disease are represented in the distribution to the left. Those patients with the disease and test values to the right of the cut-off point are TP, and those with results below the cut-off are FN; that is to say, they are wrongly identified as free from the disease. Similarly, those without the disease and with test results lower than the cut-off point are TN, and those with values above the cut-off point are FP; that is, they are wrongly identified as having the disease. The extent to which the two distributions overlap, that is the false positive and false negative rates, change when the cut-off point changes.

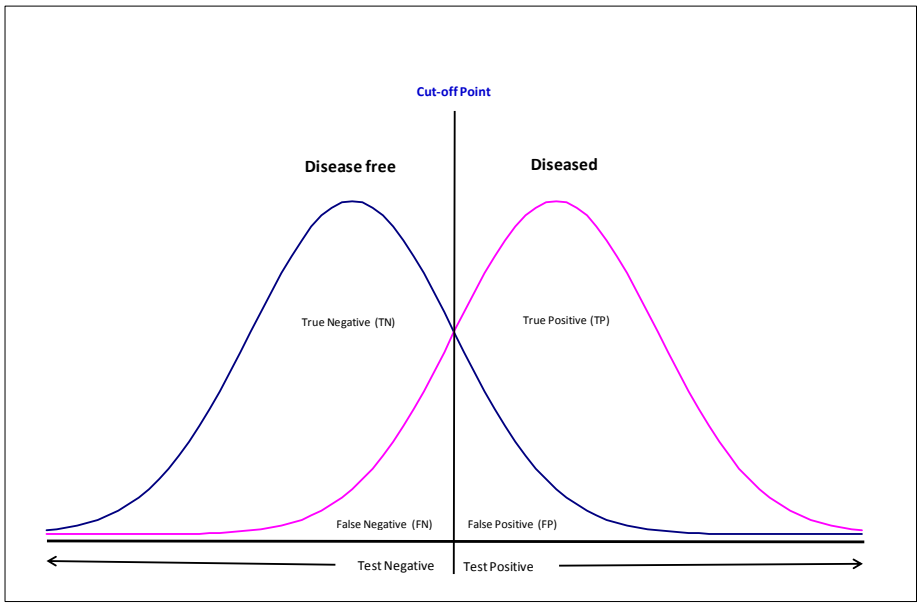


Figure 3.1 Distribution of the biomarker among disease and disease free in population

3.2.4Cut-off point

Diagnostic test results produce two types of data, qualitative data from clinical symptoms and quantitative data from diagnostic tests. The qualitative data identify patients as being with or without disease according to the presence or absence of clinical signs or symptoms. The quantitative results classify patients as diseased or disease free on the basis of whether they fall above or below the cut-off point. The cut-off point determines how many subjects are considered to have the disease. For continuous and ordinal tests, when the disease test result is above the cut-off level, the patient is assumed to have the disease. Similarly, when the test result drops below the cut-off point, the patient is assumed to be without the disease (McMaster University Health and Sciences Centre, 1981) (Edgar. et al., 1999) (Peter., 2007). Ideally, the test should not overlap in results between those with and without disease. The test would have perfect predictive accuracy and there would be no false positives or false negatives (perfect sensitivity and specificity) (Edgar. et al., 1999). A perfect test should have high sensitivity and high specificity. However, in reality most test results do not meet these standards, as in the case considered here. For most tests, the results will overlap between patients with and without disease. These relationships are demonstrated in the Figure 3.2.

Each cut-off point is related with a specific probability of true positive and false positive results. Line “A” indicates test results for high sensitivity and low specificity of about 90% and 60% respectively. All values that fall to the left are negative; those to the right are positive. This cut-off point criterion decreases the number of false negatives (increased sensitivity) but also increases the number of false positives (decreased specificity). Tests with a high sensitivity are often used to screen for disease, and tests with low sensitivity fail to identify many patients with disease. Screening tests tend to cast a wide net in order to pick up all cases of disease and not miss anyone, but they include some accidental positive results in people who do not actually have the disease. Moving the cut-off point may affect the sensitivity and specificity of the test result. Cut-off point “B” is intermediate between the two (sensitivity 80%, specificity 80%). This cut-off point shows high sensitivity and high specificity in screening tests, but still produces false positives and false negatives. Cut-off point “C” shows hypothetical test results with high specificity at about 95%, but limited sensitivity at about 60%. All values that fall to the right of “C” are considered positive; those to the left are negative. This cut-off point criterion increases the number of FN (increased specificity) but also decreases the number of FP (decreased sensitivity). Tests with a high specificity are appropriate to confirm a suspected diagnosis, but cannot be used to exclude the presence of disease because it shows low sensitivity. Patients who produce positive results on a very sensitive screening test may produce negative results on a specific confirmation test.If a test is designed to confirm a disease in

a population, a cut-off point with greater specificity and low sensitivity is selected. On the other hand, if a test is designed to screen in a general population without clinical signs or symptoms of that disease, a cut-off point with greater sensitivity and low specificity is selected.

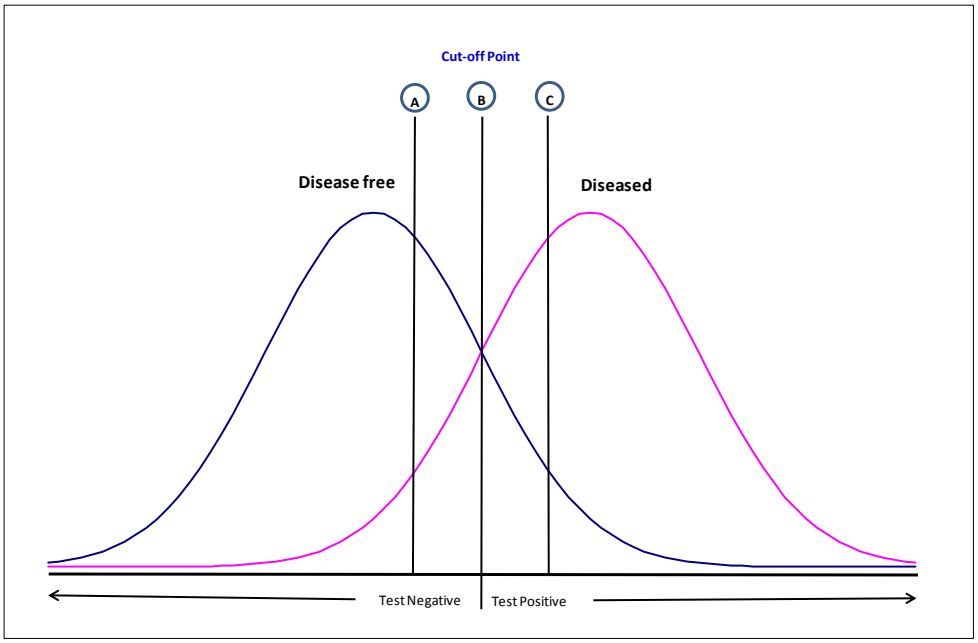


Figure 3.2 Distribution of the biomarker of results with different cut-off points

3.2.5 Test accuracy

Diagnostic accuracy relates to the ability of a test to discriminate between people with and without the disease. Sensitivity, specificity, predictive values, and likelihood are also discussed in this section. Different measurements of diagnostic accuracy relate to the different aspects of the diagnostic procedure. Some measures assess the discriminatory power of the tests; others are used to estimate the predictive ability of the test. Some results of the test relate to the background characteristics of the population to which they are applied. Sensitivity, specificity, positive predictive value and negative predictive value can all be represented as probabilities. A summary of all terms, calculation methods and definitions related to test accuracy are shown in Table 3.4 and are denoted the same as in Table 3.2.

Sensitivity and specificity

Sensitivity and specificity are two critical components that reflect the accuracy of diagnostic tests (Altman and Bland, 1994b, NCSSM Statistics Leadership Institute, 1999) (NCSSM Statistics Leadership Institute, 1999). The sensitivity of a test is the probability that a test provides a positive result when the subject does in fact have the disease, and

specificity is the probability that the test result is negative given that the individual tested is free from the disease. Sensitivity and specificity relate only to the characteristics of the test and are unaffected by population characteristics, such as the prevalence of a disease, and so can be applied to a variety of populations (Šimundić, 2008) (Altman and Bland, 1994b).

Sensitivity and specificity are of considerable importance for a clinician. Ideally, good test results should provide high sensitivity and specificity. For example, screening and diagnostic tests can cause a test to have very high sensitivity, but sometimes these test results have low specificity. The clinicians are able to keep both specificity and sensitivity high in the test, but the tests still produce FP and FN. In a large population it is impossible to avoid FP and FN. FP are especially undesirable when screening a serious disease in a population. The clinicians do not want to tell patients that they have a serious disease when they do not actually have it (NCSSM Statistics Leadership Institute, 1999).

Both sensitivity and specificity are closely related to the concepts of type I and type II errors, as shown in Table 3.3. In the ideal test, perfect test prediction can achieve 100% sensitivity, predicting that all people from the sick group are sick, and 100% specificity, not predicting that anyone from the health group is sick. The upper left corner relates to the correct decision to reject null hypothesis when the alternative is really true. The lower right corner corresponds to the correct decision not to reject the null hypothesis when it should not be rejected. A type I error is a FP result during diagnostic testing, while a test with a high specificity has a low type I error rate. The upper right corner relates to a test result that causes clinicians to reject the null hypothesis when it is actually true. A type II error is a FN result during diagnostic testing, while a test with a high sensitivity has a low type II error rate. The lower left corner corresponds to a test result that causes clinicians not to reject the null hypothesis when it is not true.

Table 3.3 Relationships among type I error and type II error with diagnostic test

	Disease	
Test	**Present (D+)**	**Absent (D-)**
Positive (T+)	True Positive (TP)	False Positive (FP) **(Type I error)**
Negative (T-)	False Negative (FN) **(Type II error)**	True Negative (TN)

Predictive values

Other statistical relationships between test results and the disease outcome are positive predictive value and negative predictive value. Positive predictive value (PPV) gives the probability that a patient has the disease given that they test positive, and negative predictive value (NPV) gives the probability of a patient not having the disease given that they test negative (Parmigiani, 2002). Unlike sensitivity and specificity, predictive values are highly dependent on the prevalence of the disease in the population. The PPV and NPV should only be used if the prevalence ratio of patients in the disease group is the same. With high prevalence the positive predictive value will be high for disease. The NPV moves in the opposite direction. For example, if a clinician uses a diagnostic test in a high prevalence population, the positive test result will be more likely to be truly positive than in a low prevalence population. Therefore, neither predictive value from one study should be applied to another setting in which prevalence differs (Šimundić, 2008) (Altman and Bland, 1994a). To overcome the problem of two populations having equal prevalence, positive and the negative likelihood ratio should be reported instead of PPV and NPV, as likelihood ratios do not depend on prevalence.

Likelihood ratio

Likelihood ratio (LR) is a very crucial measure of diagnostic accuracy. The likelihood ratio for positive test results (LR+) tell a clinician how likely it is that a patient has disease when there is a positive test result. It is usually higher than 1 because it is more likely that the positive test result will occur in subjects with the disease than in disease-free subjects. Likelihood ratio for negative test results (LR-) tell a clinician how much more likely a negative result is to be found in subjects without disease than in subjects with disease. It is usually less than 1 because it is less likely that a negative test result will occur in subjects with disease than in subjects without disease (Greenhalgh, 1997) (Šimundić, 2008). Both specificity and sensitivity are used to calculate the likelihood ratio. LR+ and LR- are unaffected by prevalence of disease. Therefore, the likelihood ratio from one study can be applied in other settings, as long as the definition of the disease is not changed (Šimundić, 2008) (Jyoti and Richard, 2009).

Prevalence

The prevalence of disease in the population can be denoted by P(D+), the prior probability of a randomly selected individual from the population having the disease.

Table 3.4 Definition of terms related to test accuracy

Term	Calculation	Definitions
Sensitivity or True positive rate (TPR)	$P(T+\mid D+) = \frac{a}{a+c}$	Sensitivity is defined as the proportion of patients with the disease who have a positive test result. This is $P(T+\mid D+)$ = true positive divided by the sum of true positive plus false negative.
False positive rate (FPR)	$P(D-\mid T+) = \frac{b}{b+d}$ 1- Specificity	The false positive rate is defined as the proportion of patients without the disease who have a positive test result. This is $P(T+\mid D-)$ = false positive divided by the sum of false positive plus true negative.
Specificity or True negative rate(TNR)	$P(T-\mid D-) = \frac{d}{b+d}$	Specificity is defined as the proportion of patients without the disease who have a negative result. This is $P(T-\mid D-)$ = true negative divided by the sum of true negative plus false positive.
False negative rate (FNR)	$P(D+\mid T-) = \frac{c}{c+a}$ 1- Sensitivity	The false negative rate is defined as the proportion of patients with disease who have a negative test result. This is $P(T+\mid D-)$ = false negative divided by the sum of true positive plus false negative
Positive predictive value (PPV) or post-test probability of a positive test	$P(D+\mid T+) = \frac{a}{a+b}$	The positive predictive value is defined as the likelihood of patients with a positive test result will have the disease. This is $P(D+\mid T+)$ = true positive divided by the sum of true positive plus false positive.
Negative predictive value (NVP) or post-test probability of a negative test	$P(D-\mid T-) = \frac{d}{c+d}$	The negative predictive value is defined as the likelihood of patients with a negative test result will not have disease. This is $P(D-\mid T-)$ = true negative divided by the sum of true negative plus false negative.
Positive likelihood ratio	$LR+ = \frac{\text{sensitivity}}{(1\text{-specificity})}$	The positive likelihood ratio is defined as the increase in the odds of having the disease after a positive test result. This is LR+ = true positive rate(sensitivity) divided by the false positive rate(1-specificity)
Negative likelihood ratio	$LR- = \frac{(1\text{-sensitivity})}{\text{specificity}}$	The negative likelihood ratio is defined the decrease in the odds of having the disease after a negative test result. This is LR- = false negative rate(1-sensitivity) divided by the true negative rate(specificity)
Accuracy	Accuracy = $\frac{a+d}{a+b+c+d}$	Accuracy is defined as the proportion of all tests that give a correct result. This is true positive plus true negative divided by the sum of true positive, true negative, false positive and false negative.
Prevalence	P (D+)	Prevalence is defined as the proportion of patients who have the disease (McMaster University Health and Sciences Centre, 1981).

The ideal test is one that has very high sensitivity and specificity, so that the most true disease cases are identified and most non disease cases are excluded. Sometimes a test result can be positive in patients who do not actually have the disease, which is called the

false positive rate (1 – sensitivity), and can be negative in patients who do actually have the disease, which is called false negative rate (1-specificity). However, sensitivity and specificity change in opposite directions when the cut-off point of tests change, this is due to a trade-off between maximising sensitivity and specificity, as mentioned in section 3.2.3. This occurs because tests generally do not have a 100% TPR or a 100% TNR. For example, as the cut-off point for positivity is high, specificity will increase and sensitivity will decrease. Diabetes is diagnosed based on a fasting blood sugar >126 mg/dl, however If a clinician moves the cut-off point to 170mg/dl, it makes it more difficult to detect positive cases. This makes the test less sensitive (some true diabetic cases don't have such high blood sugar levels) and more specific (people without diabetes may at times have blood sugar levels higher than 126mg/dl, but it is unlikely to be as high as 170 mg/dl). On the other hand, when lowering the cut-off point, the test becomes more sensitive but less specific. The score that is chosen as the cut-off point is determined by maximising sensitivity (true positive rate) and 1- specificity (false positive rate) across a series of cut-off points. Sensitivity and specificity always have this inverse relationship and the plot of the trade-off between sensitivity and the false negative rate is known as the ROC curve, which highlights the covariation between the two outcomes (Warner, 2004). The best diagnostic tests will be those that maximize both sensitivity and specificity.

If the detection or diagnosis of disease involves the use of more than one diagnostic or screening device, the evaluation of the diagnostic strategy entails a combination of two or more diagnostic and screening tests. Decision trees are an ideal tool for such evaluations. The standard decision tree for a diagnostic test begins with the disease prevalence at the first chance node and is followed by the diagnosis testing (Phelps and Mushlin, 1988). The reason why decision trees for diagnostic tests should start with disease prevalence and then be followed by the accuracy of the tests is explained at the end of section 3.3.3. Diagnostic testing is undertaken in order to identify the disease, and the accuracy of the test depends on how well the technology or test correctly indentifies the disease. The standard approach for measuring correct and incorrect identification is through diagnostic test accuracy, that is to say, sensitivity and specificity. Therefore, the test characteristics of sensitivity that can correctly identify TP and of specificity that can correctly indentify TN are important in the evaluation of diagnostic technologies. Moreover, neither sensitivity and specificity change with the prevalence of a disease, unlike PPV and NPV (Warner, 2004).

3.3 Bayesian methods for test accuracy

This section will introduce an application of Bayes' Rule in the evaluation of diagnostic tests, and show how Bayes' Rule can be used to switch between conditional and unconditional test accuracy.

3.3.1 Fundamentals of Bayesian

Bayes' theorem, is a logical consequence of the product rule of probability (Khamis, 1990). The theory of conditional probability and Bayes' theorem are found in various applications in formulating mathematical models in all sciences (Hardeo, 1992). Bayes' rule in medical diagnosis is applied and discussed in various texts, including Lusted (1968) and Sox et al (1998) (Parmigiani, 2002). Furthermore, to evaluate laboratory tests and the principles and techniques of medical decision analysis, Bayes' rule is commonly applied.

Bayes' Rule is a way of calculating conditional probabilities. Bayes was the first to use probability theory inductively, which developed the mathematical basis for probability inference (Lesaffre et al., 2007). The concept of conditional probability provides information about how the occurrence of one event predicts the probability of another event. The essential fundamentals of the Bayesian method are their estimated unknown probability, and making decisions on the basis of new (sample) information (Okeh and Ugwu, 2008). In other words, Bayesian data analysis is a practical method for making inferences from data, using probability models for quantities we observe and for quantities about which we wish to learn.

3.3.2 Simple statement of theorem

Conditional probability is a very helpful method and is used in many ways. Bayes' theorem associates the conditional and marginal probabilities of event A and B. Bayes' theorem in this form indicates a mathematical representation of how the conditional probability of event A given B is associated with the converse conditional probability of B given A. We symbolize conditionality by using a vertical slash '$|$' , which can be referred to as ' given'.

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

Each term in Bayes' theorem states:

P(A|B)	Shows the conditional probability of A given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
P(B|A)	States the conditional probability of B given A. It is also called the likelihood.
P(A)	Indicates the prior probability or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
P(B)	Indicates the prior or marginal probability of B, and acts as a normalizing constant.

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

Each term in Bayes' theorem states:

P(B|A)	Shows the conditional probability of B given A. It is also called the posterior probability because it is derived from or depends upon the specified value of A.
P(A|B)	States the conditional probability of A given B. It is also called the likelihood.
P(B)	Indicates the prior probability or marginal probability of B. It is "prior" in the sense that it does not take into account any information about A.
P(A)	Indicates the prior or marginal probability of A, and acts as a normalizing constant.

3.3.3 An application of Bayes' Rule to screening and diagnostic test evaluation

The function of a diagnostic test is to make a diagnosis. Clinicians have to know the probability that the test results will give the correct diagnosis. In fact, medical diagnosis tests commonly yield mathematic test results such as sensitivity and specificity. However, the sensitivity and specificity of test results alone do not provide strong enough guidance

to make clinical decisions. They cannot be converted into clinically relevant quantities without information on disease prevalence. On the other hand, probabilities related to screening and diagnosis tests are based on Bayes' theorem, and is applied to diagnostic test evaluation. Bayes' rule shows that positive and negative predictive values can be calculated from the sensitivity and specificity of diagnostic tests, which are values that a clinician needs in order to make appropriate diagnoses or decisions. (Khamis, 1990) (Anthony, 2007) (Okeh and Ugwu, 2008). Suppose you test positive for a disease. What is the probability that you actually have the disease? It depends on the accuracy and sensitivity of the test, and on the background (prior) probability of the disease. Let "D+" stand for "disease present" and "D-" stand for "disease absent". Let "T" indicate test results and "+" the event of a positive test and "-" the event of a negative test. In particular, the prevalence rate in the population is represented as P(D+), the sensitivity as P(T+ | D+) and the specificity as P(T- | D-).

There are two possible decision tree maps which involve diagnostic tests. Firstly, the initial branching of the decision tree starts with testing. According to a decision tree map, a patient visiting the clinic undergoes tests and examinations to distinguish whether the patient is with or without disease. Treatment or further tests will be offered according to those test results. The final results represent whether disease is present or absent in the patient. Such, Bayes theory uses conditional probability to assess decision making under uncertainty and is particularly applicable to decision trees. Bayes' rule is therefore applied to determine the probability of a particular diagnosis, given the appearance of specific signs, symptoms and test outcomes (MedicineNet, 2004). Figure 3.3 presents a common example involving one diagnostic test where treatment is given following the test result. The beginning of the decision tree map starts with receiving a positive or negative test result. The probabilities are denoted by P(T+) and P(T-). The disease outcomes must be shown in the right hand of the test results as the disease status. The subsequent branching probabilities are the likelihood of disease in each outcome. The probabilities are denoted by P(D+ | T+), P(D- | T+), P(D+ | T-), and P(D- | T-).

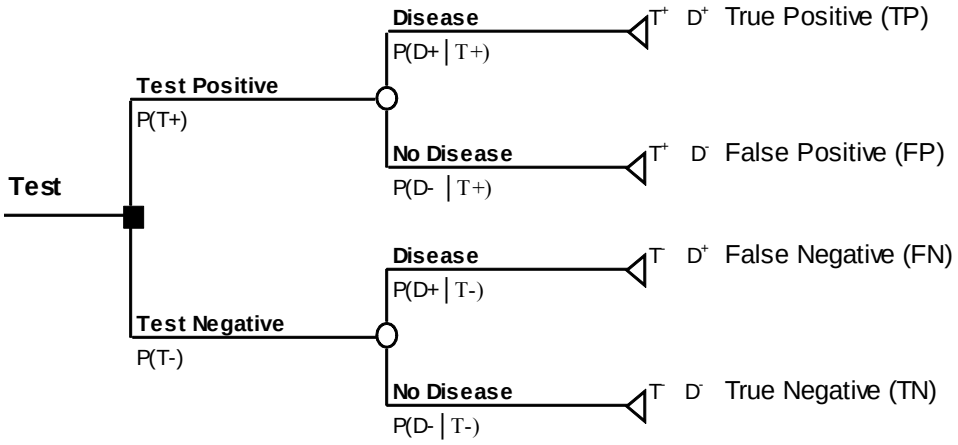


Figure 3.3 Tests approach to decision problem with testing

The four possible final outcomes associated with this decision tree are then calculated

True Positive	=	P(T+) * P(D+ | T+)
False Negative	=	P(T+) * P(D- | T+)
False Positive	=	P(T-) * P(D+ | T-)
True Negative	=	P(T-) * P(D- | T-)

We can compute the posterior probability P(D+ | T+) of a patient who tested positive while actually having the disease. In this context, Bayes' rule takes the from

$$\text{Positive predictive value} = P(D+ \mid T+) = \frac{P(T+ \mid D+)P(D+)}{P(T+)}$$

$$\text{Positive predictive value} = P(D+ \mid T+) = \frac{P(T+ \mid D+)P(D+)}{P(T+ \mid D+)P(D+)+P(T+ \mid D-)P(D-)}$$

$$\text{Positive Predictive Value} = \frac{(\text{Sensitivity})(\text{Prevalence})}{(\text{Sensitivity})(\text{Prevalence})+(\text{1-Specificity})(\text{1-Prevalence})}$$

P(D+)	The probability that the disease is present in the patient, regardless of any other information. This is the prior probability of D.
P(T+)	The probability of a positive event, which is found by adding the probability that a true positive result will appear with the probability that a false positive will

appear.

P(T+ | D+) This is the probability of true positive rate, that is, that the test is positive and the disease present.

P(T+ | D-) This is the probability of false positive rate, that is, that the test is positive, though the disease is absent. The prior probability of P(T+ | D-) is 1 - P(T+ | D+).

Similarly, we can also compute the posterior probability P(D- | T-) of a patient who tests negative while actually being without the disease, from Bayes's theorem.

$$\text{Negative predictive value} = P(D- \mid T-) = \frac{P(T- \mid D-)P(D-)}{P(T-)}$$

$$\text{Negative predictive value} = P(D- \mid T-) = \frac{P(T- \mid D-)P(D-)}{P(T- \mid D-)P(D-)+P(T- \mid D+)P(D+)}$$

$$\text{Negative Predictive Value} = \frac{(\text{Specificity})(1\text{-Prevalence})}{(\text{Sepcificity})(1\text{-Prevalence})+(1\text{-Sencitivity})(\text{Prevalence})}$$

P(D-) The probability that the disease is absent in the patient. The prior probability of D- is 1- P(D).

P(T-) The probability of a negative event, which is found by adding the probability that a true negative result will appear with the probability that a false negative will appear.

P (T- | D-) This is the probability of true negative rate, that is, that the test is negative and the disease absent.

P(T- | D+) This is the probability of false negative rate, that is, that the test is negative, though the disease is present. The prior probability of P(T- | D+) is 1 - P(T- | D-).

Secondly, the other method of constructing the decision tree commences with branches of disease status. The subsequent branching states the test outcomes. The primary branching probability shows the likelihood of outcomes for the patient within the specific population, in terms of presence or absence of disease. In each branch, the incidence or prevalence within the specific population are indicated. The probabilities are denoted by P(D+) and P(D-). The test outcomes must be shown in the right hand branch of the

disease status. The subsequent branching probabilities are the likelihood of tests in each outcome. The probabilities are denoted by P(T+ | D+), P(T+ | D-), P(T- | D+), and P(T- | D-). In particular, subsequent pathways for the probabilities of sensitivity and specificity are indicated. These probabilities are indicated by the test results, which are routinely calculated and reported, as shown in Figure 3.4.

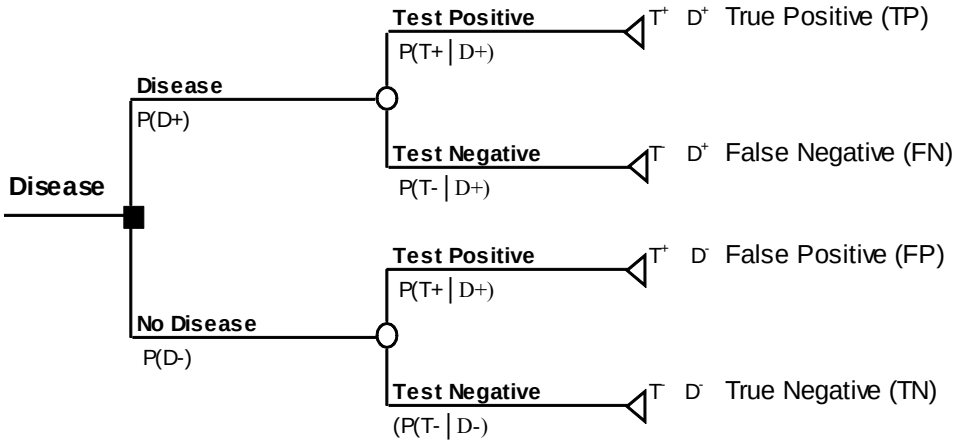


Figure 3.4 Test approaches to decision problems with disease status

The four possible outcomes associated with this decision tree are than calculated:

True Positive	=	P(D+) * P(T+ | D+)
False Negative	=	P(D+) * P(T- | D+)
False Positive	=	P(D-) * P(T+ | D-)
True Negative	=	P(D-) * P(T- | D-)

The actual chorological method used in the second method of constructing the decision tree, whereby the initial chance node separates the population according to the disease status involves the same information as the first method, and it can be illustrated that the probabilities associated with each pathway are equivalent in the two methods.

True positive from the first method $= P(T+)*P(D+|T+)$

$$= P(T+)*\frac{P(T+|D+)*P(D+)}{P(T+)}$$

$$= P(D+|T+)*P(D+)$$

True positive from the second method = $P(D+|T+)*P(D+)$

$\therefore$ True positive from first method = true positive from second method

Whilst it is not possible to indentify the actual disease status for an individual patient consulting with the doctor, using prevalence and incident data it is possible to calculate the expected numbers with and without disease for a given population. Hence, when modelling a health policy decision affecting a population, the second approach can be used, enabling the direct application of conditional probabilities to the following branches. Given that most medical information that is provided is conditional upon the disease present and absent, the second method provides an easier method by which to examine a population based decision problem. In light of this, applications of decision analysis to economic evaluations of health care commonly employ the second method.

3.4 Combinations of the test results

Combinations of the tests are usually applied to many diagnostic, health-certification, and disease-surveillance situations. Multiple tests might be applied to all subjects or the tests might be used on a subset of a population. Decision rules are then used to test the results and to classify individuals as disease positive or negative (Gardner et al., 2000). In these cases, clinicians can combine two component tests under either the “negative rule” or “negative dominant strategy” and either the “positive rule” or “positive dominant strategy” (Tang, 2004). Positive and negative dominant strategies may help the clinician to interpret the combination of test results. Before introducing the positive and negative dominant strategies, discussions of the differences in characteristics of screening tests (Test 1) and diagnostic tests (Test 2) which divide people into the two groups are provided. Normally, in order to classify disease in a population, a two step approach is used. The population is initially screened by means of a screening test. Patients that test positive for the screening test undergo a diagnostic test to confirm the results (Public Health Action Support Team 2010). In a two step approach, the initial clinical performance of the laboratory is shown as four outcomes, namely TP, FN, FP and TN, as introduced in the previous section. Similarly, diagnostic tests are shown as four outcomes, according to the test results.

The following section illustrates an example using two binary tests to explain the concept of negative and positive dominant strategies. There is an example with a basic branch of a decision tree for intervention A to represent the dominant strategy, as shown in Figure 3.5 and Figure 3.6. The decision tree initially starts with the probability of disease being present (prevalence) or absent (1 – prevalence). Branches of decision nodes present possible outcomes of probability for the prevalence. Subsequently, each point

where there is a chance node indicates the screening test (Test 1) and diagnostic test (Test 2) strategies. Branches of the chance nodes illustrate the possible positive (T+) or negative (T-) test results. Those branches apply the probability of sensitivity and specificity of screening and diagnostic tests. The diagnostic test accuracy (sensitivity and specificity) for Test 1 is initially applied indentifying TP (sensitivity 1) and TN (specificity 1) for correct diagnoses, whereas FN (1 – sensitivity 1) and FP (1- specificity 1) are for incorrect diagnoses. Following test 1, test 2 is performed and likewise the diagnostic test accuracy for test 2 is applied identifying TP (sensitivity 2), FN (1 – sensitivity 2), FP (1 – specificity 2) and TN (1 – specificity 2).

3.4.1 Negative dominant strategy

The negative dominant strategy (NDS) represents a test strategy in which negative test results dominate. Figure 3.5 demonstrates an example of a decision tree for the NDS. A two step approach that uses screening test and diagnostic tests is used to demonstrate the NDS. In the negative dominant strategy, if screening tests (Test 1) show negative results, patients will not receive a second test to confirm the results. On the other hand, diagnostic tests (Test 2) are used to confirm the results if the screening tests (Test 1) are positive. NDS accepts positive outcomes over negative results. For example, positive results (TP) are only obtained if Test 1 and Test 2 are both positive. If either Test 1 or Test 2 is negative a negative result is obtained. There are a greater proportion of TN and FN outcomes in the NDS where screening and diagnostic tests involve negative results. Therefore, NDS benefits from the specificity of both tests, at the expense of lower overall sensitivity, and a greater proportion of FN identifications. Positive test results are dominated by negative test results, which are significant as all patients who have disease but test negative, are likely to not receive treatment or have delayed treatment. Moreover, in order to calculate the sensitivity and specificity of the strategy, the values for the sensitivity and specificity of each test can be combined using the formulae below. The sensitivity of Test 1 and Test 2 are denoted by Sen_t1 and Sen_t2, respectively, and the specificity of both tests are denoted by Spe_t1 and Spe_t2, respectively.

$$\text{Sensitivity} = \text{Sen\_t1}*\text{Sen\_t2}$$

$$\text{Specificity} = \text{Spe\_t1} + \text{Spe\_t2} - (\text{Sen\_t1}*\text{Spe\_t2})$$

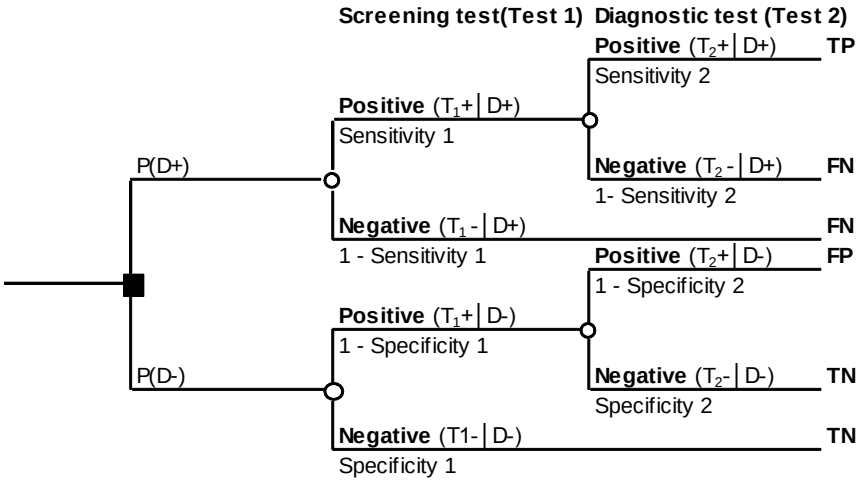


Figure 3.5 Decision tree for the negative dominant strategy (NDS)

3.4.2 Positive dominant strategy

The positive dominant strategy (PDS) represents a test strategy in which positive test results dominate. As the two step approach demonstrates, in the P(D+) pathway if a screening test (Test 1) is over the threshold all patients who test positive directly receive treatment without further testing, as shown in Figure 3.6. In the other pathway P(D+), patients that are negative for the screening test (Test 1) undergo an additional diagnostic test (Test 2). In other words, PDS would accept positive outcomes over negative results. For example, if a patient has a negative screening test result and a positive diagnostic test the strategy treats this as a positive result. In contrast, only if the results of both Test 1 and Test 2 are negative will the outcome for the strategy be treated as negative. Consequently positive test results in the PDS dominate negative test results. Sensitivity of both of the tests benefits PDS, and may result in a larger proportion of TP being identified than could be done through either test alone, this is because in this strategy FN results from Test 1 are countered by positive outcomes in Test 2. However the PDS presents with few FN and a larger proportion of FP, which means trading-off specificity for greater sensitivity. In terms of patient management, PDS may result in more FP which is likely to lead to over-diagnoses. It means that people who do not have the disease may be treated unnecessarily. Moreover, to calculate the sensitivity and specificity of the strategy, as in the previous section, a formula that combines the values for each test can be used.

Sensitivity = Sen_t1+Sen_t2-(Spe_t1*Sen_t2)

Specificity = Spe_t1*Spe_t2

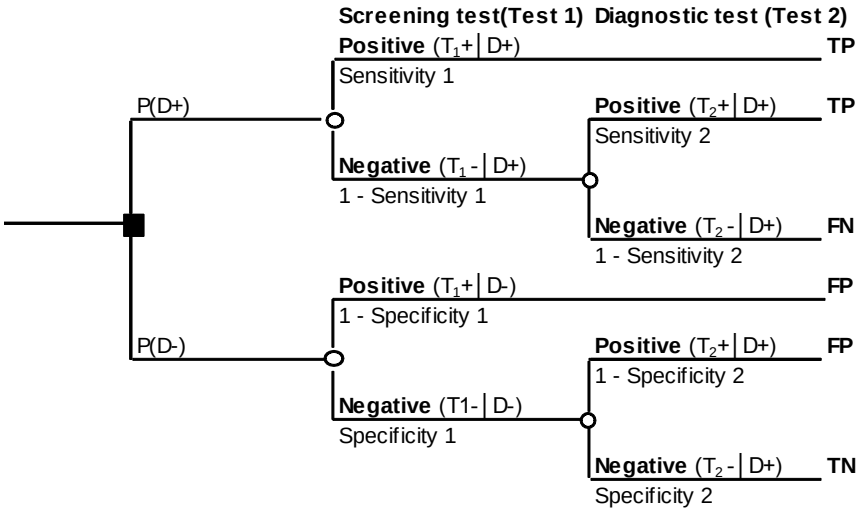


Figure 3.6 Decision tree for the positive dominant strategy (PDS)

However, most diagnostic technologies have adopted NDS. All screening and diagnostic test devices classify patients with some error, clinicians must select from a variety of methods to interpret the test results, trading-off the risk of FP and FN (Phelps and Mushlin, 1988). The NDS and PDS approaches allow the clinician to consider test results in terms of the differences in FN and FP test results. The trade-off between sensitivity and specificity or detection and unnecessary testing is at the heart of screening and diagnostic tests and the accuracy of the tests in terms of sensitivity and specificity is an important consideration for clinicians.

Moreover, combining the tests, NDS and PDS give alternative sensitivity and specificity trade offs compared to individual tests. In individual tests, even if clinicians are able to keep both specificity and sensitivity high, FP and FN outcomes will still persist. The individual tests involve diagnostic tests without prior screening. Consequently, the accuracy of the individual tests, in terms of sensitivity and specificity, always have this inverse relationship, and the plot of the trade-off between sensitivity and the false negative rate presents as the optimal point in the ROC curve. However, on combining the tests, when the second test is performed to confirm the first test, there is a trade-off between the sensitivity and specificity of the tests. The two-step approach it consider better than using individual tests.

3.4.3 Positive and negative dominant strategies in various screening test approaches

Screening and diagnostic tests are medical procedures used to detect a disease in individuals in the general population without clinical signs or symptoms of that disease.

Screening and diagnostic tests involve controversial issues that include who should be screened, whether a one-step or two-step approach should be employed and whether screening should be universal or selective. This section discusses how to apply NDS and PDS to all screening and diagnostic approaches, as shown in Figure 3.7.

Treatment options

There are two different treatment decisions for screening tests outlined in Figure 3.7, whether to treat all patients or not to treat all patients. If none of the patients are given treatment by the clinician, all classifications are either TN or FN for treatment. Patients without disease that do not receive treatment classify as TN (negative for the receipt of treatment and true as the patients do not require treatment due to the lack of disease). Likewise, patients with disease that have not received treatment identify as FN (negative for the receipt of treatment and false as the patients should have received treatment due to the presence of disease).

If on the other hand the whole population receives treatment irrespective of disease state, all patients are then classified as either TP or FP. In this instance TP is indicative of patients with disease that received treatment and FP represents patients without disease that receive treatment.

One step approach option

The one step approach is a diagnostic test performed without prior screening. Results for both positive and negative dominant strategies are listed. For both strategies however, patients that test positive classify as TP whereas those that test negative display as FN. Likewise patients without disease that test positive present as FN, while TN is representative of patients without disease who test negative.

Two step approach option

In this approach, a diagnostic test is performed to confirm the results of a positive screening test. Outcomes of both positive and negative dominant strategies for the two step approach are similar to those discussed in sections 3.4.1 and 3.4.2.

Universal screening option

Universal screening refers to the screening of all pregnant women in the population and so this strategy ignores the risk factor pathway at the beginning of the decision tree. In this

option, positive and negative dominant strategies are applied to universal screening tests for GDM. It is assumed that all patients receive both screening tests (Test 1) and diagnostic tests (Test 2).

For the positive dominant strategy, patients with disease that test positive for either Test 1 or Test 2 are classified as TP, and those that are negative in both Test 1 and Test 2 present as FN. For patients that do not have disease but test positive in either Test 1 or Test 2, the results are FP, whereas those that test positive in both Test 1 and Test 2 are TN.

On the other hand, with regard to the negative dominant strategy, patients with disease that test positive in both Test 1 and Test 2 have TP outcomes, while a negative result in either Test 1 or Test 2 gives a FN outcome. For patients without disease, those that have positive results for both Test 1 and Test 2 are FP, while those with either a negative result in Test 1 or Test 2 are TN.

Selective screening option

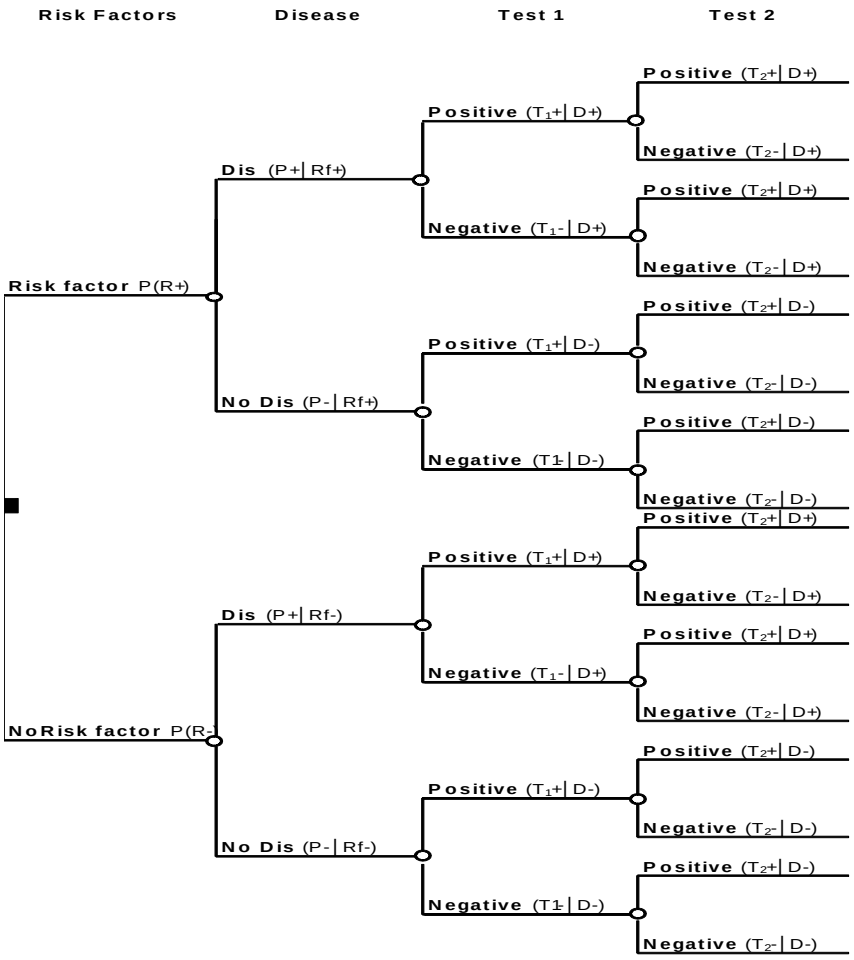
This strategy assumes that at first visit, all patients are only screened for risk factors. In this model risk factor screening is considered to be a selective screening tool. Patients with risk factors may present with disease or without disease. For patients with risk factors, those that have disease are TP, whereas patients who are disease absent classify as FP. On the other hand, for patients without risk factors, if disease is present they are TN and if absent they are FN.

Selective screening and other two tests options

As in the previous strategy, all patients are screened for risk factors at the first visit, however in this option all women that test positive for risk factors receive further tests. The outcomes for both positive and negative dominant strategies, combining the risk factor screening test and Tests 1 and Test 2, are listed in Figure 3.7.

For patients that are positive for risk factors in the PDS, if they are positive for either Test 1 or Test 2 and have disease the outcome is TP for risk factors, whereas patients that do not have disease and are negative in either Test 1 or Test 2 are FN for risk factors. On the other hand for patients that are negative for risk factors, but have disease and are positive for either Test 1 or Test 2, the outcome is FP for risk factors. When all three tests are negative and the patient has no disease the outcome is TN.

In the NDS, if all tests are positive and disease is present, the outcome is TP. Additionally for patients that test positive for risk factors, if there is no disease or for those that have disease and either Test 1 or Test 2 is negative the outcome is FN. In the group with no risk factors, for those patients that have disease and who test positive in both Test 1 and Test 2, the outcome is FP. Also for patients without disease or those that have disease but are negative for either Test 1 or Test 2 the outcome is TN.



Treatment		Universal		One step approach		Two step approach		Risk factor	Risk factor +test	
No	All	PD	ND	PD	ND	PD	ND		PD	ND
FN	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP
FN	TP	TP	FN	TP	TP	TP	FN	TP	TP	FN
FN	TP	TP	FN	FN	FN	TP	FN	TP	TP	FN
FN	TP	FN	FN	FN	FN	FN	FN	TP	TP	FN
TN	FP	FP	FP	FP	FP	FP	FP	FP	TP	FN
TN	FP	FP	TN	FP	FP	FP	TN	FP	TP	FN
TN	FP	FP	TN	TN	TN	FP	TN	FP	TP	FN
TN	FP	TN	TN	TN	TN	TN	TN	FP	FN	FN
FN	TP	TP	TP	TP	TP	TP	TP	FN	FP	FP
FN	TP	TP	FN	TP	TP	TP	FN	FN	FP	TN
FN	TP	TP	FN	FN	FN	TP	FN	FN	FP	TN
FN	TP	FN	FN	FN	FN	FN	FN	FN	FP	TN
TN	FP	FP	FP	FP	FP	FP	FP	TN	FP	TN
TN	FP	FP	TN	FP	FP	FP	TN	TN	FP	TN
TN	FP	FP	TN	TN	TN	FP	TN	TN	FP	TN
TN	FP	TN	TN	TN	TN	TN	TN	TN	TN	TN

Dis = Disease; PD = Postive dominate; ND = Negative dominate: TP = True positive; FP = False positive: FN = False negative; TN = Treu negative

Figure 3.7 The different strategies of screening test for GDM

3.5 The economics of diagnosis

Economic assessments of diagnostic tests are necessary for health technology assessments (HTA) (Sanghera et al., 2013). There are however two main reasons why the economic assessment of diagnostic tests can be difficult. Firstly, difficulties can arise because of doubts about the relationship between diagnosis and health outcomes. Secondly, because of the increasing importance of diagnostic technology in healthcare, only with economic assessments will the most value be gained from restricted medical resources (Mushlin et al., 2001). Diagnostic technologies make it possible for clinicians to make decisions that allocate the right therapy to the right patient. The outcomes of diagnostics tests are more usually measured in terms of case detection and the sensitivity and specificity of test results. To translate this information into an economic evaluation it is necessary to perform a cost-effectiveness analysis (CEA). Moreover, new diagnostic technologies could potentially incur substantial additional expense, creating the need for comparative economic and clinical analysis.

3.6 Conclusion

Screening and diagnostic tests relate to the ability of a test to discriminate between the target condition and health. In other words, they identify between the presence or absence of disease in patients. The discriminative potential of disease can be quantified by the measure of diagnostic accuracy. Diagnostic accuracy can be measured in different ways, such as sensitivity, specificity, predictive values and likelihood ratios. To use the results of test accuracy, a clinician should consider the prevalence in the population, as positive and negative predictive values are highly dependent on the disease prevalence in the population. Bayes' rule is a useful theorem, which can be applied in the diagnostician's inference. It illustrates the sensitivity and specificity of diagnostic tests and can calculate positive and negative predictive values, which are quantities that a clinician needs in order to make appropriate diagnoses or decisions. Laboratory tests are imperfect and may identify some healthy people as diseased (false positive result) or identify some diseased patients as disease-free (false negative result). Although screening tests and diagnostic tests are a critical contributor to clinical decision making, test results may have unwanted and unintended consequences. Test results may interfere with clinical decisions if the tests poorly identify between people with and without disease. All tests must be chosen properly and used with deliberation and purpose with the expectation that the test results will decrease ambiguity surrounding patient problems and contribute to their health.

Chapter 4 Gestational Diabetes Mellitus (GDM)

4.1 Introduction

In developing an economic evaluation, it is important to fully understand the disease including how the disease presents itself, how the disease changes over time and how treatment affects the disease. The purpose of this chapter is to provide a review of the literature regarding Gestational diabetes mellitus (GDM). The chapter is divided into 6 main sections and begins with a definition of the disease area. In the next section, a brief review is given of the short and long term maternal and perinatal complications. At the end of this section, the burden to both the health care system and finances will be demonstrated. Next, the clinical management of GDM are described where treatments fall into two main groups, namely pharmacological and non-pharmacological. In the following section, numerous guidelines, which have been developed and published in Europe and UK, are described by year, testing, threshold values and type of screening test. After this, the various screening and diagnostic tests in current use, based on these guidelines, are then described. In the last section, an account is then given of different screening and diagnostic test guidelines in current use and newly published in the UK and Scotland.

4.2 Search strategy

The aim of this chapter is to provide a literature review of screening tests for GDM. The following electronic databases were searched from 1996 to December 2013: Medline, Embase, and Web of Science. Specific searches were performed covering five main areas (screening test, burden of disease, guidelines, management or treatment, adverse complications) for each of the databases, as detailed in Appendix 1. Inclusion criteria were applied to include relevant publications that provided information on screening tests for GDM in terms of the management or treatment of GDM, burden of disease, guidelines, adverse complications, diagnostic tests for GDM, quality of life related to GDM and type 2 DM. In addition, the prevalence of GDM and economic evaluations of screening tests of GDM are presented in a systematic review in Chapter 5 and Chapter 6, respectively and the search strategies of those systematic reviews are therefore not included in the results of this Chapter. The search outputs are detailed in Figure 4.1. After assessment, 104 studies remained for final review. Information from the review was then used to design the economic evaluation model to identify appropriate comparators, and parameter estimates for the model.

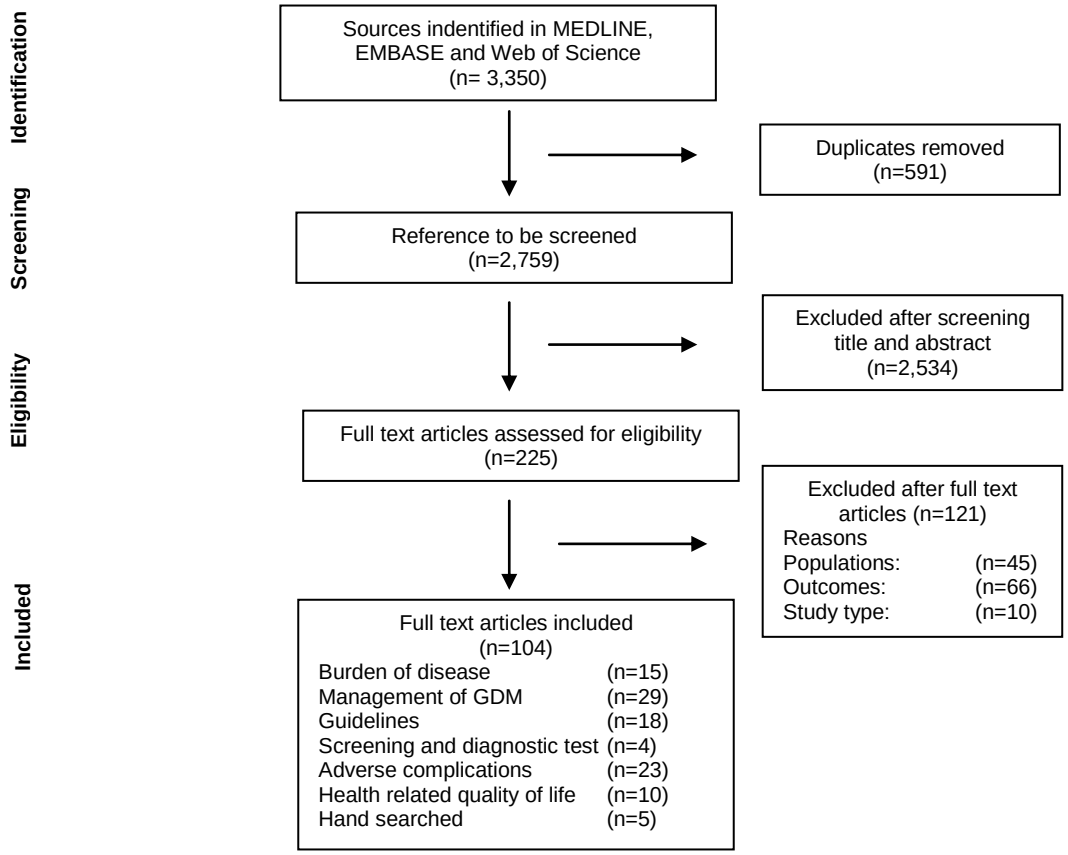


Figure 4.1 Flowchart of article selection

4.3 The disease

Diabetes mellitus (DM) is a disease that has been recognised for thousands of years (Sattley, 2008). There are two major forms of maternal diabetes which may occur during pregnancy: GDM and pre-existing diabetes (NHS., 2011). Firstly, GDM is the most common medical complication of the pregnancy period. In 1999, the World Health Organisation (WHO) defined GDM "as carbohydrate intolerance of variable severity with onset or first recognition during pregnancy, whether or not insulin is used, and regardless of whether diabetes persists after pregnancy". Women who become pregnant and who are known to have DM which antedates pregnancy do not have gestational diabetes but have "diabetes mellitus and pregnancy" and should be treated accordingly before and after the pregnancy (WHO, 1999a). GDM is a relatively new condition that has only been recognized in modern times, since the nineteenth century. In the 1940 – 1950s it was recognized that women developing type 2 DM had excess perinatal mortality and that their babies were born with excessive weight (Kitzmiller, 2010). Secondly, GDM is different in pregnant women with pre-existing diabetes, including mothers that have type 1 and type 2 diabetes mellitus before pregnancy. Type 1 DM is marked by the body's inability to produce insulin on the one hand, and type 2 DM is actually caused by the body's resistance to insulin (NHS., 2011). Care and treatment for

these must be different from that for GDM. Women who become pregnant and who have either type 1 and type 2 DM before pregnancy are not said to have GDM, but are said to have pre-existing diabetes mellitus, and should be treated before, during and after the gestation period (American Diabetes Association, 2006) (NHS., 2011). GDM resolves after delivery in approximately 90% of the cases reported by Kjos and colleagues (Kjos et al., 1990). Patients who had GDM in previous pregnancies may also get GDM again in subsequent pregnancies.

4.3.1 Pathophysiology

GDM is a particular type of diabetes developed by some women during pregnancy. GDM is a condition in which the patient suffers from glucose intolerance as it occurs in patients with type I and type II diabetes. Pregnant women, in particular, are vulnerable to transient glucose intolerance because, during pregnancy, the hormones make it harder for the body to respond to insulin. The metabolism is changed in normal pregnancy to ensure adequate nutrition of the foetus. Levels of maternal estrogen and progesterone hormones increase at the beginning of pregnancy, and promote pancreatic β-cell hyperplasia and increased insulin release. GDM is a group of diseases that are caused by deficiencies in insulin secretion and/or insulin action that occurs when pancreatic function is not sufficient to overcome the insulin resistance created by a change in diabetogenic hormones during pregnancy (Jovanovic, 2010). As pregnancy is characterised by insulin resistance and hyperinsulinemia, it may predispose some women to develop diabetes. In pregnancy, the placenta normally produces diabetogenic hormones, such as growth hormone, corticotrophin-releasing hormone, placental lactogen and progesterone which induce maternal insulin resistance and compensatory hyperinsulinemia. Pregnancy hormones and other factors result in interference with the action of the insulin receptor. These and other endocrinologic and metabolic changes ensure that the foetus has an adequate supply of energy and nutrients at all times. In some pregnancies, pancreatic insulin production cannot compensate for this insulin resistance, and carbohydrate intolerance develops (Kennedy et al., 2006). Two forms of insulin resistance occur in women who develop GDM. The first of these is the physiological insulin resistance of late pregnancy. The second is a chronic form that is present before pregnancy and becomes worse because of the physiological changes that exacerbate insulin resistance during pregnancy (Metzger et al., 2007).

The placenta is a foetal organ with widespread functions that contains a common boundary between mother and foetus. The functions of the placenta are transport of maternal nutrients to sustain foetal growth, synthesis of hormones and growth factors to facilitate maternal adaptation to pregnancy, provision of an immunologic barrier, and

dissipation of heat energy from foetal metabolism (Hiden and Desoye, 2010). Women with GDM have elevated blood glucose levels in spite of higher insulin levels. Pancreatic secretion of insulin is increased, but insulin does not reduce blood glucose levels. Subsequently, extra blood glucose passes through the placenta, resulting in high foetal levels. This, in turn, causes the baby's pancreas to produce extra insulin to eliminate the blood glucose. Since the infant is getting more energy than it needs to grow and develop, the extra energy is stored as fat (Jones, 2001).

4.3.2 Symptoms

DM commonly presents during pregnancy because the process of gestation in the non-diabetic women makes significant demands on insulin-producing cells (Fox and Pickering, 1995). Most mothers with GDM do not have any symptoms and the condition is detected by screening. Sometimes, pregnant women with GDM may have symptoms of high blood sugar, including increased thirst, polyuria and tiredness. However, those are also common symptoms of pregnancy.

4.4 Burden of gestational diabetes mellitus

This section presents a review of the published literature on burden of GDM. The burden of GDM can be categorised into both health and financial aspects. The major health burden of GDM is the prevalence of maternal and infant morbidity. On the other hand, the financial burden is very much the cost of screening and treatments in order to complete full diagnostic and medical procedures required for both the mother and infant.

4.4.1 Health burden

Reports from many sources indicate that type 2 DM is increasing rapidly throughout developed and developing countries. For example, Hilary King and co-workers have carefully analysed a number of reports and published their findings, in which adjustments for age were included, to develop the most comprehensive estimates of the prevalence of diabetes (King et al., 1998). In the same study, predictions were made of the numbers of people with diabetes who would be aged 20 years or more in three different time periods, 1995, 2000 and 2025. Worldwide, the prevalence of diabetes in adults was estimated to be 4.0% of the population (135 million adults) in 1995 and it is estimated that in 2025 the prevalence of diabetic adults will increase to 5.4% of the population (300 million adults) In 2004, the revised projections of the prevalence of diabetes for year 2000 and 2030 were published by Wild and colleagues. The prevalence of diabetes for all age groups worldwide in 2000 and 2030 was estimated at

2.8% and 4.4% respectively. The total number of people with diabetes is predicted to increase from 171 million in 2000 to 366 million by 2030 (Wild et al., 2004). Moreover, Type 2 diabetes and obesity are frequently diagnosed in children and young adults in many countries. In the UK, 2.6 million people have been diagnosed with diabetes mellitus in 2009. The prevalence of diabetes in adults across the UK varies, from 5.1 % in England, 4.5 % in Northern Ireland, 4.6 % in Wales and 3.9 % in Scotland (Diabetes UK, 2010).

The global increase in incidence of diabetes has been accompanied by an increase in incidence of GDM in many countries. There are differences in the results of prevalence studies. The prevalence of women with GDM has increased over time, affecting between 1%-16% of women, depending on the diagnostic criteria and population studies (King, 1998). A recent international survey of 47 countries estimated a prevalence range of <1% - 28%, with data derived from expert analysis and national prevalence estimates (Jiwani et al., 2012). Moreover, GDM has been found to be more prevalent in African Americans, Hispanic/Latino and American Indians (National Institute of Diabetes and Digestive and Kidney Diseases, 2008). The prevalence is often significantly different in populations of different size and diversity, reflecting differences in geographic diversity (states, regions, and countries) and the definition used to identify women with GDM. The magnitude of the risk varied in different ethnic groups, ranging from 9% in Caucasians, 11.9% in Latinos, and 25% in women of Mediterranean or east-Asian descent (Berger et al., 2002). Moreover, approximately 650,000 women give birth in England and Wales each year, and 2-5% of them have diabetes (NICE, 2008b). As mentioned above, the prevalence of GDM has not yet been fully quantified. Additionally, the prevalence of pregnancy with GDM has increased over time. This study therefore conducted a comprehensive systematic review to assess the prevalence of GDM and is presented in chapter 5.

Increased maternal morbidity because of GDM may occur during pregnancy or in the longer term (Barry and Gabbe, 1998). In many studies, an increased risk was reported for preeclampsia, polyhydramnios, and cesarean section in women with GDM (Sermer et al., 1995) (de Veciana et al., 1995). Despite the fact that GDM occurs during pregnancy, a relatively short period of time in a woman's life, it may cause type 2 DM in the long term. Postpartum, women with GDM have a significantly increased risk of type 2 diabetes. The National Institute of diabetes and digestive and kidney disease in 2008 stated that the risk of developing diabetes in women with GDM was about 5-10%, and that the risk slightly increased at 40%-60% in the next 5 to 10 years. Coustan and colleagues studied former gestational diabetic women, and found diabetes or impaired glucose tolerance (IGT) in 6% at 0–2 years, 13% at 3-4 years, 15% at 5-6 years, and

30% at 7-10 years postpartum (Coustan et al., 1993). In the case of postpartum impaired glucose tolerance and high body mass index (BMI) in adult females, they are predicted to develop type 2 diabetes after GDM occurrence.

Previous estimates of perinatal mortality have been based on results of older studies. Evidence to support increases in perinatal mortality related to GDM was confirmed in a more recent literature review of these older studies (Martine et al., 2007). Perinatal mortality rates of 49-198 per 1,000 births have been observed to occur in women receiving GDM treatments (Coustan and Lewis, 1978). The perinatal mortality rate in infants of diabetic mothers has declined sharply from 250 per 1,000 live births in 1960 to a near-normal 20 per 1,000 live births in 1980 (Weintrob et al., 1996). Infants born to diabetic mothers have a risk of perinatal morbidity and mortality, resulting from hyperbilirubinemia, macrosomia, birth trauma, hypoglycaemia and neonatal respiratory distress syndrome (Crowther et al., 2005).

In terms of health-related quality of life (HRQoL) in women with GDM, there have been only a few studies that have investigated the effects of GDM on women's HRQoL. A randomised control trial (RCT) in Australia measured maternal health status by the Short-Form 36 (SF-36) at 6 weeks and 3 months postpartum, and stated that treatment of GDM improved the mother's HRQoL (Crowther et al., 2005). The results of Crowther and colleagues were used to estimate QALY in a cost-effectiveness analysis for GDM in the UK (Round et al., 2011). In Finland, the HRQoL after pregnancy of 100 women sampled from the birth register at a University hospital was measured using the 15D instrument. This study showed insignificant difference in the median and mean level values of 15D between the GDM group and the control group (Halkoaho et al., 2010). In addition, another report showed that Type 2 DM after GDM may reduce life expectancy up to 10 years on average (Diabetes UK, 2012). No studies have been conducted in terms of Disability adjusted life years (DALY) for GDM yet. However, two cost-effectiveness analysis (CEA) studies for GDM screening tests reported results in terms of DALY Adverse in order to present the long term adverse complications for mothers who have both GDM and type 2 DM (Lohse et al., 2011) (Marseille et al., 2013).

4.4.2 Financial burden

It is estimated that 10% of the entire National Health Service budget is accounted for by diabetes, approximately £9 billion a year (based on the 2007/2008 budget for the NHS) (Diabetes UK, 2010). GDM management involves initial screening and diagnostic tests, treatments (dietary therapy, self-monitoring blood glucose level, pharmacotherapy and insulin programme), management of maternal medical complications (maternal trauma,

preeclampsia and operative deliveries), monitoring blood glucose level (pharmacotherapy and insulin programme), and management of neonatal complications (macrosomia, brachial plexus injury (BPI), jaundice and birth trauma). All of these account for parts of the financial budget.

In the UK, the NHS reported national costs for antenatal care and diabetes in pregnancy in 2008 (NICE, 2008a). The cost of a random blood test was £3.37 and the cost of a diagnostic test by 75g oral glucose tolerance test was £17.58. It was estimated that the net costs of screening and testing for GDM in England were £2,150,000 per year, which included all pregnant women with risk factors tested by the biochemical test and diagnostic test. The estimated costs of treatment of blood glucose monitoring, oral medication and regular insulin or analogue insulin were £704,000, £4,000 and £55,000 respectively. The costs of implementation of treatment for GDM by year 1, 2 and 3 in England were estimated at £840,000, £897,000 and £953,000, respectively (National Health Service, 2008b). In 2007, Chen and colleagues estimated the national medical costs associated with GDM by analysing National Hospital Discharge data in the USA. They showed that the total estimated cost attributable to GDM nationally was $596 million (approximately $3,305 per woman with GDM) for mothers, $40 million for newborn babies (approximately $209 per newborn of mothers with GDM), and $320 million for medical care costs (Chen et al., 2009). Maternal GDM can be managed in various ways. Approximately 65% (cost £114.82 per course/person) of the cases will undergo dietary therapy with constant monitoring of blood glucose levels, approximately 20% (cost £3.09 per course/person) will receive oral hypoglycaemia therapy, and approximately 15% (cost £59.57 per course/person) will be treated with insulin (National health service, 2008a). In 2007, The University Hospitals of Coventry and Warwickshire estimated that the cost of each case of OGTT to the NHS was £12.13 (Wilson et al., 2008).

4.5 Adverse complications of GDM for mother and offspring

GDM is associated with substantial increases in rates of maternal and prenatal complications. During gestation and after delivery, both mother and offspring are at risk of developing short and long term complications. In a meta-analysis conducted by Marine and colleagues in 2007, strong associations were reported between GDM and adverse outcomes for mother and children including hypertensive disorder, caesarean section, macrosomia, infant respiratory distress syndrome and hypoglycemia (Martine et al., 2007).

When a mother has metabolic aberrations during pregnancy complicated by diabetes, there are serious implications for the foetus, newborn baby, child and young adult (Weintrob et al., 1996). According to the “Pederson hypothesis”, maternal diabetes is characterized by high concentrations of glucose that result in increased nutrient transfer to the foetus. Maternal glucose crosses to the placenta easily, which results in intrauterine hyperglycemia, foetal hyperinsulinemia and possible modification of growth and development of the foetus. (Pedersen, 1954). GDM during pregnancy can have multiple adverse outcomes for the mother and offspring, which can be divided into short and long term, shown in Table 4.1. This review will present these adverse outcomes by dividing them into short and long term impacts.

Table 4.1 Short and long term complications in mothers and offspring

	Age at Expression	Period of Exposure	Complications
Mother with GDM			
Short term	Gestation	3rd trimester	Hypertension, Preeclampsia
		Delivery	Cesarean section, Instrumental delivery (forceps and ventouse deliveries), Vaginal injuries or Uterine hemorrhage
Long term	After delivery		Obesity , GDM in next gestation, type 2 DM
Offspring in mother with GDM			
Short term	Fetus	2nd trimester	Macrosomia, Nervous system development delay
		3rd trimester	Chronic hypoxemia, Stillbirth
	Newborn	Deliver	Birth injury, Respiratory distress syndrome, Polycythemia and Hyperbilirubinemia, Neonatal metabolic problems (Hypoglycemia, Hypocalcemia, Hypomagnesemia ,Hyperbilirubinemia)
Long term	Child/adult		Obesity, Impaired glucose tolerance, DM

The details of short and long term complications in both mothers and offspring in Table 4.1 have been simplified and shown in diagrammatic form in Figure 4.2. The figure is divided into 4 sections. The upper two sections show short term complications in both mothers and offspring. In mothers, hypertension and preeclampsia are presented as direct complications which occur during the gestation period. Whereas, cesarean section, instrument deliveries and vaginal injuries are presented as indirect adverse complications which occur in mothers who have macrosomic babies, shown by broken arrow lines in the figure. Moreover, offspring with macrosomia can produce indirect complications not only for mothers but also for themselves. Long term adverse complications in both mothers and offspring are shown in the two lower sections of this figure. Obesity and type 2 DM are the main long term complications.

The design of the diagram in Figure 4.2 of short and long term complications does not taken into account perinatal mortality. Perinatal mortality refers to the death of a foetus or neonate and is considered to be the most important complication of GDM. Martine and co-workers found in their meta-analysis that the main evidence for an increased risk of perinatal mortality associated with GDM stemmed only from older

studies (Martine et al., 2007). Additionally, in a current systematic review and meta-analysis of the effects of treatment in women with GDM, zero cases of perinatal mortal were reported. Because of this, pinatal mortality is not taken into account in the presentation of short term complications in Figure 4.2.

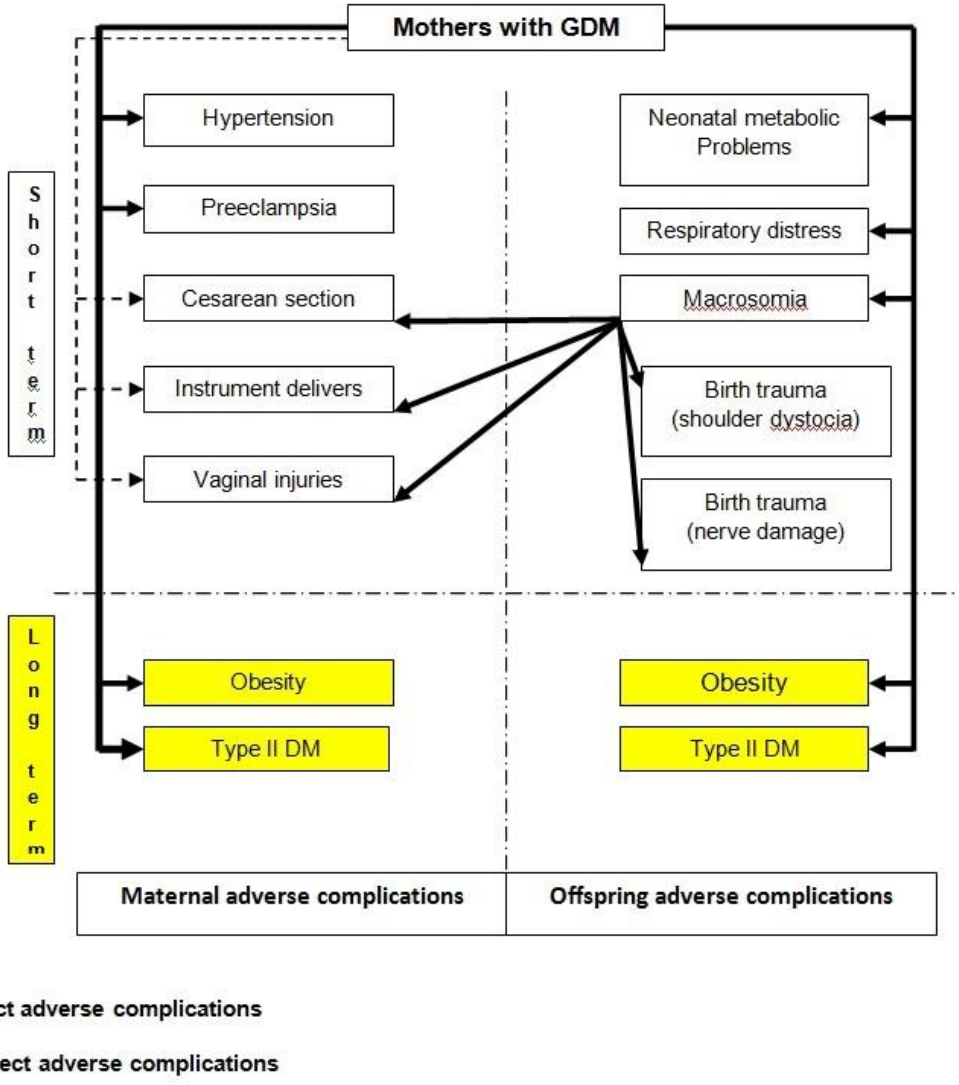


Figure 4.2 Short and long term complications in both mother and offspring

4.6 Management of gestational diabetes mellitus

The goal of treatment in women with GDM is to achieve blood glucose level control. Diabetic control in pregnancy and obstetric management will be associated with a reduction in the severity of neonatal complications and improve the adverse outcomes for the mother. In terms of the management of GDM, effective treatment demonstrates that diagnosis and treatment of GDM results in a reduction in adverse effects for both mother and child. In two randomized control trials, it has been shown that interventions

in women diagnosed with GDM using dietary advice, monitoring and management of blood glucose, are effective in reducing birth weight, the rate of growth for gestational age infants, and perinatal morbidity. The authors of the Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS), a large randomized control trial (RCT) of treatment for GDM, concluded that treatment of GDM reduced the risk of serious perinatal complications and may also improve women's health-related quality of life during pregnancy and after delivery (Crowther et al., 2005). The authors of another RCT in 2009, The Maternal Foetal Medicine Units network (MFMU) study, found that treatment of mild GDM did not significantly reduce the frequency of a composite outcome that included stillbirth or perinatal death and several neonatal complications but did reduce the risk of foetal overgrowth, shoulder dystocia, caesarean section, and hypertensive disorders (Landon et al., 2009). In the ACHOIS, a significant decreased risk in the composite outcome was demonstrated, but in the MFMU study the risk of composite outcome was not significantly different. The authors of a systematic review of treatment of women with GDM stated that treatment for GDM, including treatment to control blood glucose alone or with other specific treatment, seems to lower the risk for some perinatal complications in both mother and child (Horvath et al., 2010). In addition, untreated carbohydrate intolerance during pregnancy was associated with increased risk of maternal and perinatal morbidity (Pettitt et al., 1980). Bellamy and colleagues found that proper tests and management such as dietary, lifestyle and pharmacological intervention, might prevent or delay the onset of type 2 diabetes in women with GDM (Bellamy et al., 2009).

There are a substantial number of treatments for GDM. This review divides these into two main groups, non-pharmacological and pharmacological, based on the Scottish Intercollegiate Guidelines Network for the management of diabetes (SIGN) and the National Institute for Health and Clinical Excellent (NICE) guidelines for antenatal care diabetes in pregnancy (SIGN, 2010) (NICE, 2008d). Management of GDM in these guidelines are the same, including dietary therapy and blood glucose monitoring on the one hand, and pharmacotherapy (Glibenclamide and Metformin) or insulin therapy (Regular insulin, Insulin aspart and Insulin lispro).

4.6.1 Dietary therapy

Dietary therapy is the primary GDM treatment strategy for controlling glucose levels and ensuring appropriate maternal weight gain. A good diet is important for women during pregnancy, as food provides the mother's energy requirement and supplies the foetus with the nutrients it needs to develop. All women should receive nutritional advice from an appropriately skilled dietician. For women with diabetes before pregnancy, pre-

pregnancy nutritional assessment and care from a dietician may improve the patient's chance of achieving the best possible diabetes control before conception. Dietary programmes need to conform to the principles of dietary management of diabetes in general and meet the nutritional requirements of pregnancy. In addition, dietary therapy should be individually tailored for each woman with GDM and take account of maternal weight and body mass index. The dietary recommendations generally include avoiding sugar, convenience foods and junk foods. Small frequent meals and a very small breakfast should be eaten and fat intake reduced. Generally, daily calorie needs for women of normal weight in the second half of pregnancy are 30 – 32 kcal per kilogram of body weight. Dietary therapy plans control blood sugar by controlling the amount of carbohydrate food consumed during a day. Eating less carbohydrate may decrease blood sugar level. Some carbohydrates are digested more slowly and raise blood sugar level more than others (Diabetes prevention and control program, 2007).

Carbohydrate counting for meals and snacks are considered in most dietary programme recommendations. The American Diabetic Association (ADA) recommends an intake of calories and nutrients sufficient to decrease maternal hyperglycemia. The dietary programme recommended a 30-33% reduction of calories for those pregnant with diabetes, with at least 40% daily carbohydrate intake. The ideal allocation would be 50-60% carbohydrates (complex and high fiber), 10 -20 % protein, and 25 - 30 % fat (< 10% saturated). In 2003, the British Diabetic Association recommended not less than 45% carbohydrates with a low glycaemic index, and a limited proportion of dietary fats. The Fifth Workshop Conference on GDM in Chicago in November 2005 suggested an optimal carbohydrate content in pregnancy of 175 g/day including 140g for the mother and 35g for foetal brain metabolism (Cianni D G et al., 2008).

Major and co-workers found that women with GDM who received dietary treatment and who obtained less than 40 percent of their total calories from carbohydrates, had babies with lower birth weights and had lower risk of caesarean delivery than women with higher intakes(Major et al., 1998). Magee and colleagues, who performed a randomized control trial in obese women with GDM, stated that restriction of carbohydrates by 50% (1,200 kcal/day) improved glycaemic control but, significantly increased the risk of ketonemia and ketonuria (Magee et al., 1990). The authors of a systematic review of the evidence for screening for GDM found that more than 70% of patients with GDM have mild hyperglycemia and are usually treated with diet alone (Brody et al., 2003).

4.6.2 Blood glucose Monitoring

Once patients with GDM start a dietary programme, it is important to monitor capillary blood glucose levels. Patients and health care providers have to increase the frequency of the monitoring of daily blood glucose levels. Monitoring the blood glucose is important in evaluating the efficacy of the exercise and eating programme by testing blood sugar levels at the proper times during each day. In the self-monitoring of blood glucose, a drop of blood is tested for blood glucose levels by using a special device. Patients should do this test three times a day, as shown in Table 4.2. Occasionally, some patients may ask to test more frequently during a day or at night.

The Australasian Diabetes in Pregnancy Society (ADIPS) recommends self-monitoring of blood glucose, obtaining at least one fasting and one 1 or 2 hour postprandial (5.5, 8.0 and 7.8 mmol/L respectively) tests daily (Hoffman et al., 1998). The Fourth International Workshop Conference on Gestational Diabetes Mellitus suggested maintaining fasting blood glucose concentrations at less than 5.3 mmol/L, and for one or two hours after a meal at less than 7.8 and 6.7 mmol/L respectively, glucose levels which are lower than ADIPS (Metzger and Coustan, 1998). The clinical practice recommendation of the American Diabetes Association (ADA) on GDM is to perform fasting glucose 5.8 mmol/L and 2-hour postprandial plasma glucose 6.7 mmol/L tests. Self-monitoring of blood glucose in glycaemic control is a controversial issue (American Diabetes Association, 1997). Several studies had shown that more frequent self-monitoring improved glycaemic control. Postprandial hyperglycemia is more closely related to fetal macrosomia than preprandial hyperglycemia in pregnancy with diabetes. In one randomised study of postprandial and preprandial blood glucose monitoring in women with GDM who required insulin treatment, patients who measured their glucose levels after a meal had infants with lower birth weight, and had fewer caesarean deliveries (de Veciana et al., 1995).

Table 4.2 Treatment targets for women with gestational diabetes mellitus

Test	Glucose levels mmol/L		
	The American Diabetes Association	The Australasian Diabetes in Pregnancy Society (ADIPS)	The Fourth International Workshop Conference on Gestational diabetes mellitus
Fasting capillary (venous plasma)	< 5.8	< 5.5	< 5.3
1 hour postprandial capillary (venous plasma)	-	< 8.0	< 7.8
2 hour postprandial capillary (venous plasma)	< 6.7	< 7.8	< 6.7

4.6.3 Insulin therapy

Insulin therapy is used when capillary blood glucose levels exceed the threshold, according to the guidelines for self-monitoring blood glucose levels, because of a lack of effectiveness of dietary therapy in controlling these levels. The decision regarding starting insulin therapy differs greatly from centre to centre and is based on many factors, including effectiveness of screening, local incidence, mode of surveillance, individual obstetrical, medical and psychosocial factors, and local resources (Firth, 1996).

Various criteria have been proposed for starting insulin therapy in women with GDM. The American College of Obstetricians and Gynaecologists Committee (ACOG) recommends insulin therapy for women whose fasting glucose level is over 5.2 mmol/L, whose one-hour postprandial glucose level is over 7.1 to 7.7 mmol/L, or whose two-hour postprandial glucose level is over 6.6 mmol/L (ACOG 2001). The ADA recommends a higher glucose level for starting insulin therapy: between 5.0 to 5.5 mmol/L in the fasting state, less than 7.7 mmol/L at one hour after eating, and less than 6.65 to 7.05 mmol/l at two hours after eating (Metzger et al., 2007). Buchanan and colleagues recommend initial insulin therapy to reduce risk of macrosomia when abdominal circumference (AC) measured by ultrasound is higher than the 75th percentile for gestational age (Buchanan et al., 1994).

Insulin is the first-line pharmacologic therapy for GDM. Human insulin has several theoretical and practical advantages for GDM. It does not cross the placenta and is highly effective. In gestational diabetes, mothers require insulin for the duration of

the pregnancy period (Firth, 1996). Regular insulin, which is used in pregnancy for treatment of diabetes, has some limitations: it becomes effective from 30 – 60 minutes after subcutaneous injection, the effect peaks after 2 – 3 hours and the duration of the effect is too great (about 8 – 10 hours). Currently available insulin analogues include rapid-action mealtime insulins Lispro (Humalog) and Aspart (Novolog), and long-acting basal insulin Glargine (Lantus). Mealtime insulins (Lispro, Aspart) are used to control post-meal blood glucose level. For controlling between-meal and overnight blood glucose levels a basal insulin is used (NPH, Glargine, Lente and Ultralente). Lispro and Aspart are categorised as class B drugs in The United States of America. However, ACOG and the ADA have not yet officially recommended their use. In contrast to Lispro and Aspart, there is little information on the use of the long-acting insulin analogues Glargine (Lantus) and Detemir (Levemir) in pregnancy. HPN is the intermediate-acting option of choice for those who require pharmacologic treatment.

Initially, the insulin dose can be calculated on the basis of the patient's weight. The first doses depend on gestational age, capillary blood glucose monitoring levels, current body weight at the start of therapy, further adjustments, base activity, meal plan and other factors, as shown in Table 4.3. Most gestational diabetic patients require an insulin dose of 0.6 U/kg body weight or more. Patients may safely be given an initial total daily dose of 0.4 U/kg by giving two-thirds of it in the morning and one-third in the evening (Firth, 1996). The normal dose for diabetes patients is 0.8 U/kg, and 0.9 -1.0 U/kg for overweight and obese women.

Table 4.3 Total suggested daily insulin during pregnancy

Gestational weeks	Total daily insulin
Week 1-18	0.7 U/kg actual body weight
Week 18 – 26	0.8 U/kg actual body weight
Week 26 – 36	0.9 U/kg actual body weight
Week 36 – 40	1.0 U/kg actual body weight

In most, but not all, prospective trials of insulin therapy in women with GDM, a reduction has been shown in the risk of neonatal macrosomia (Turok et al., 2003). Coustan and Lewis reported that treatment of women with diabetes with insulin (20 units NPH and 10 Units regular) was effective in reducing the incidence of foetal macrosomia (Coustan and Lewis, 1978). Similarly, the authors of a study in Alabama in 1990 using the same insulin dose reported a successfully reduced mean birth weight, macrosomia rate, and ponderal index (Thompson et al., 1990). In contrast, little effect on birth weight, birth trauma, operative delivery, or neonatal metabolic disorder was found in a randomised controlled trial of intensive treatment of GDM (Garner et al., 1997).

4.6.4 Pharmacotherapy

Oral hypoglycaemic agents are not currently used in treating GDM because these medications cross the placenta and could stimulate the foetal pancreas. For this reason, tolbutamide, chloropropamide, and other sulfonylureas are not used in pregnancy (Langer, 1993) (Hoffman et al., 1998). Nevertheless, glyburide therapy is an alternative for women who are unable or unwilling to take insulin, and is recommended in several practices as a first-line therapy (Serlin and Lash, 2009). Glyburide, an oral glucose lowering agent often used in the treatment for type 2 Diabetes mellitus, has been shown not to cross the placenta and has been studied for use in pregnant women (Elliott et al., 1994). Moore stated that the safety of glyburide is difficult to establish because of the relatively small number of patients with GDM in the study (Moore, 2007). 201 glyburide-treated women were compared to 203 insulin-treated women. The study showed gyburide to be a clinically effective alternative to insulin therapy in women with GDM (Langer et al., 2000). In a secondary analysis of a previous paper by Langer and co-workers, it was also found that glyburide and insulin are equally efficient in the treatment of GDM at all levels of disease severity (Langer et al., 2005). In addition, metformin (Glucophage) may be another option for treatment of GDM. In a randomised controlled trial of 751 women with diabetes in urban obstetrical hospitals in New Zealand and Australia, in which treatment with metformin (plus insulin, if needed) and treatment with insulin alone were compared, neither were associated with an increase in perinatal complications (Rowan et al., 2008). In another randomised study, metformin was compared with insulin treatment in patients with GDM. Rowan and colleagues reported that glycaemic control in patients with GDM treated with metfomin and/or insulin was strongly related to pregnancy outcomes (Rowan et al., 2010).

During pregnancy, the NICE and SIGN guidelines recommend that women with GDM receive interventions that include dietary control, home glucose monitoring and pharmacotherapy, as well as insulin therapy when needed (SIGN, 2010) (NICE, 2008d). Pharmacotherapy or insulin should be considered when dietary therapy results are ineffective in controlling blood glucose levels, when there is a lack of expected weight loss following calorie restriction, or when patients are consistently hungry (Serlin and Lash, 2009). GDM treatment should be focused on the prevention of both mother and foetal complications (Kjos and Buchanan, 1999). A team focused treatment for managing GDM in women is ideal, if possible. The teams would usually include an obstetrician, diabetes physician, a diabetes educator (diabetes midwifery educator), dietician, midwife and paediatrician. The number of maternal care visits per mother is significant in confirming whether GDM is being managed properly. Bryson and colleague stated that where the number of maternity clinic visits were 80% less than expected,

women with GDM were found to have a higher risk of both eclampsia (convulsions from high blood pressure) and preeclampsia (swelling of kidney problems from high blood pressure) than women with GDM who complete prenatal care (Bryson et al., 2003). Women with GDM should be closely monitored until the early postpartum period.

4.7 Guidelines: screening and diagnostic test thresholds for GDM

The purpose of screening for GDM as early as possible is to diagnose and treat pregnant women with GDM (Martine et al., 2007). There are three main reasons for screening for GDM. Firstly, GDM is a significant complication and metabolic disorder of pregnancy, and disappears after delivery (Koukkou et al., 1995) (Carpenter and Coustan, 1982). Secondly, GDM is associated with increased risk of adverse effects for both mother and child. These include foetal macrosomia and perinatal mortality in children, hypertension disorders, pre-eclampsia, and caesarean section in mothers, and future diabetes in both (Persson and Hanson, 1998) (Greco et al., 1994). Thirdly, identifying this group of women may not only prevent perinatal morbidity but also reduce the risk of complications through diet and lifestyle management, oral medication programmes and insulin therapy (Jovanovic, 1998). In addition, Kristina and colleagues stated that unrecognized GDM increases risks of large gestational-age infants, macrosomia, shoulder dystocia and birth trauma independent of maternal obesity (Adams et al., 1998).

Numerous screening test guidelines for GDM have been developed and published. Most countries have their own diabetes associations, these societies often publish guidelines for GDM, which may differ slightly (National Diabetes Data Group, 1979) (American Diabetes Association, 1998) (WHO, 1999a) (Metzger and Coustan, 1998). The screening test strategies for GDM are still debated, and no consensus has been established yet. (Metzger and Coustan, 1998) Agreeing on the diagnostic threshold for GDM remains problematic (Ryan, 2011). A summary of recommendations from reviewed guidelines of the tests and criteria commonly used to diagnose GDM ordered by year of recommendation are shown in Appendix 2. In this section, the guidelines focus on guidelines published in Europe and the UK, as shown in Table 4.4.

In 1991, The European Association for the Study of Diabetes (EASD) first published the guidelines for GDM by using fasting plasma glucose for screening tests, 50g GCT and 75g OGTT for screening (Lind and Phillips, 1991).The most used oral glucose tolerance test internationally was a 2-hour 75g glucose solution test recommended by World Health Organization (WHO) in 1998 (WHO, 1999a). The 75g

OGTT involves one-step and two-step approaches to screening and diagnosis for GDM in all pregnant women during 24 – 28 weeks of gestation. All pregnant women have random or fasting plasma glucose tests followed by a 75g OGTT.

In 2001 The Management of Diabetes National Clinical Guidelines were published by The Scottish Intercollegiate Guidelines Network (SIGN 55) (SIGN, 2001). SIGN 55 endorsed using universal screening for GDM. In the same year, The Clinical Resource Efficiency Support Team (CREST) published the Management of Diabetes in Pregnancy. CREST is a team of health care professionals in Northern Ireland, whose guidelines recommend universal screening. The screening test suggested was RBG at 28 weeks of gestation and the recommended diagnostic test was 75g OGTT, using the same threshold as EADS in 1991(Clinical Resource Efficiency Support Team, 2001).

In 2002, Health Technology Assessment (HTA) UK concluded that while some women with very high levels of glucose should be treated, there was uncertainty over the benefits of screening across the population. 50g GCT is used in these guidelines by selective screening (Scott et al., 2002).

In 2008, the National Institute for Health and Clinical Excellence (NICE) published the Diabetes in Pregnancy Clinical Guideline 63. NICE is the relevant decision-making agency in England & Wales, independent of the government. They recommended that the 75g OGTT should be used to screen and diagnose using the criteria defined by The World Health Organization (NICE, 2008d). Subsequently, the March 2010 SIGN guidelines publication recommend a widening of diagnostic criteria and an increase in glucose tolerance testing in pregnancy (SIGN, 2010). Additional new consensus criteria for GDM reported in 2010 recommended screening tests using 75 g OGTT which based on HAPO study as outlined Diabetes Care (Moses, 2010) (Metzger et al., 2010a).

Table 4.4 Guidelines and recommendations for screening tests for GDM

Organisation	Year	Diagnostic method	Threshold values (mmol/l) 0h	1h	2h	3h	Type of screening	Remark
EADS	1991	FPG	≥4.8				Universal	
		50g GCT		≥8.2				
		75g OGTT	≥6.0		≥9.0			
WHO	1998	50g GCT		≥7.8			Universal	One-step approach to screening
		75g OGTT	≥7.0		≥7.8			and diagnosis, one or more criteria must be met or exceeded and one step approach
SIGN	2001	RPG	≥5.5				Universal	Urine for screening for every visit
		FPG	≥5.5					
		75g OGTT	≥7.0		≥9.0			
CREST	2001	RPG	≥5.5				Universal	Urine for screening for every visit and 28 weeks gestation
		75g OGTT	≥6.0		≥9.0			
HTA UK	2002	50g GCT		≥7.8			Selective	Very selective screening based on age, obesity and ethnic origin
NICE	2008	FPG	≥7.0				Selective	Women with on of risk factor should be offer screening test
		75g OGTT	≥7.0		≥7.8			
Consensus	2010	75g OGTT	≥7.0		≥7.8		Universal	
SIGN	2010	FPG	≥5.1				Selective	one or more value must be met or exceeded
		75g OGTT	≥10.0		≥8.5			

Organisation (C&C) Indicated Carpenter and Constant, (WHO) World Health Organisation, (EASD) European Association for the Study of Diabetes. (SIGN) The Scottish Intercollegiate Guidelines Network, (SOGC) The Society of Obstetricians and Gynaecologists of Canada (NICE) The National Institute for Health and Clinical Excellence : **Tests** (OGTT) Oral Glucose tolerance Test, (GCT) Glucose challenge test, (FPG) Fasting Plasma Glucose, (RPG) Random Plasma Glucose : **Threshold** Now, if you want to convert mg/dl of glucose to mmol/l, you can divide the result in mg/dl by 18 or multiply by 0.055 but, if you want to convert mmol/l of glucose to mg/dl, just multiply by 18.

In summary, over the past decade, many tests and various thresholds for GDM have been developed, not only across America, but also within Europe and the UK. The OGTT remains the foundation for the diagnosis of GDM. All expert panels such as the WHO, NICE, EADS, and SIGN recommend OGTT as the “gold standard” for the diagnosis of diabetes in pregnancy. The 100g OGTT is popular In North America, while the 75g OGTT version is used more in Europe and UK. In the UK, two national surveys of UK obstetric units have shown a lack of consensus on screening tests for GDM, with some health care centres using the ADA guidelines and others implementing the WHO guidelines. Moreover, screening and diagnostic tests, cut-off values, timings, and subsequent management programmes vary widely in the UK (Mires et al., 1999) (Hanna et al., 2008). The national survey of the Association of British Clinical Diabetologists (ABCD) reported 82 percent of centres in the UK provide routine screening for GDM, half of those centres using universal screening and half selective screening. This survey reports a widely varied prevalence of GDM from 0.1-10% (Median 1.5%) (Hanna et al., 2008). Consequently, the prevalence of GDM is significantly related to the screening test used in its detection and the cut-off point.

4.8 Guidelines: screening and diagnostic test thresholds for postpartum screening for type 2 DM in women with a history of GDM.

Many guidelines suggest postpartum follow up glucose tolerance tests for mothers with GDM (WHO, 1999a) (ACOG, 2000) (American Diabetes, 2004) (NICE, 2008d) (SIGN, 2010). The WHO recommends postpartum screening for type 2 DM by 75g OGTT at 6 weeks or more after delivery but does not recommend it for follow-up after this time (WHO, 1999a). The ACOG proposed that the postpartum screening for type 2 DM should be performed at the time of the postpartum visit by FPG or 75g OGTT. According to the ADA guidelines, GDM mothers should have their blood glucose tested by RPG or FPG within 1-3 days post-delivery (American Diabetes, 2004). If the test is negative, the next test should be delayed until 6-12 weeks after delivery in GDM women who do not have DM immediately postpartum. Then, mothers should receive a confirmation test again around their early postpartum visit using 75g OGTT. If glucose levels are normal on all previous follow ups, reassessment of glycemia should be undertaken at a minimum of 3 year intervals. NICE proposed using RPG (but not an OGTT) at 6 weeks postpartum and annually thereafter (NICE, 2008d). The variable recommendations for postpartum screening and continued monitoring for type 2 DM in women with history of GDM is shown in Table 4.5.

Table 4.5 Guidelines and Recommendations for postpartum screening for type 2 DM in women with history of GDM

Organisation	Time	Diagnostic method	Threshold values (mmol/l)				Remarks
			0h	1h	2h	3h	
ADA	Post-delivery (1-3 days)	RPG FPG	≥5.2	≥9.9	≥8.5	≥7.7	
	Early postpartum visit	75g OGTT	≥6.0		≥9.0		Two or more value must be met or exceeded
	1 year postpartum	75g OGTT	≥6.0		≥9.0		Two or more value must be met or exceeded
	Tri-annually	75g OGTT	≥6.0		≥9.0		Two or more value must be met or exceeded
WHO	6 weeks or more after delivery	FPG 75g OGTT	≥6.1 ≥6.0	≥9.9	≥11.1 ≥9.0	≥7.7	
ACOG	No specific recommendation, but should be the first postpartum visit	75g OGTT	≥6.0		≥9.0		FPG can be used in subsequent testing if both FPG and OGTT are normal postpartum
NICE	1 year postpartum 2 year postpartum 3 year postpartum	RPG RPG RPG	≥5.1 ≥5.1 ≥5.1				Offer lifestyle advice
SING	6 weeks postpartum	FPG	≥5.1				With 75g OGTT if clinically indicated

The postpartum screening for type 2 DM is a controversial issue in terms of the appropriate time and methods for screening (Bentley-Lewis et al., 2008). Most guidelines for the postpartum screening of type 2 DM in women with history of GDM

recommend screening witthin different time frames. Ketzmiller and co-workers conducted a systematic review of mostly European studies, and found the prevalence of type 2 DM, 1-10 years after delivery, to be 2.3% – 9.3% (Kitzmiller et al., 2007). Kim and colleagues estimated the efficacy and cost of postpartum screening for diabetes among women with GDM. Screening tests for diabetes with FPG, OGTT, HbA1c, annually, every 2 years and every 3 years over a period of 12 years were simulated and compared and the in the outcomes of the study it was stated that screening women with a history of GDM using OGTT every 3 years had the lowest costs per case of detected diabetes (Kim et al., 2007b). Therefore, an appropriate time frame for postpartum screening should be longer than 6 years but not shorter than 10 years to mitigate against the risk developing type 2 DM and it is recommended that an annual test using OGTT be repeated every 2 or 3 years.

4.9 Screening and diagnostic methods for gestational diabetes mellitus

The usual test for detecting GDM is screening of all pregnant women by measurement of plasma glucose between the 24th and 28th week of gestation. (Naylor et al., 1997). In a recent survey reaching 173 countries, 47 countries responded and reported that they have GDM testing and treatment guidelines. Many countries use a variety of screening approaches, including universal screening (routine screening of all pregnant women), selective screening (based on risk factors) or a mixed approach. In this survey, the most commonly used screening and diagnostic tests were the FPG, the 50g GCT, the 75g OGTT and the 100g OGTT, dependant on the guidelines (Jiwani et al., 2012). This section presents screening procedures, screening tests and diagnostic tests based on UK and European guidelines, as mentioned in a previous section.

4.9.1 Screening procedures

Screening and diagnostic tests are a controversial topic in terms of the appropriate screening of pregnant women with GDM. (Greene, 1997) (Jarrett et al., 1997) (Soares J de AC, 1997). Other controversial issues include who should be screened, whether a one-step or two step approach should be employed and whether screening should be universal or selective.

Universal and selective screening

A screening test can be applied universally or selectively. In screening for GDM there is controversy regarding the use of universal or selective screening (Vogel et al., 2000)

(Coustan, 1991). Normally, screening and diagnostic tests are performed between 24 and 28 weeks because at this time in gestation the diabetogenic effect occurs in pregnancy. There is sufficient time during pregnancy to start treatment or therapy programmes (Greene, 1997). The Universal screening strategy is to screen all pregnant women by measurement of plasma glucose between 24 and 28 weeks using different test techniques. On the other hand, the selective screening strategy is to screen initially all pregnant women with risk factors. In the next stage, women who show positive risk factors undergo further screening. Both universal and selective screening may operate by one-step or two-step protocols, depending on the screening criteria. In the two-step protocol, women who have exceeded the threshold for screening tests have to continue with diagnostic tests. Confirmation of GDM is made diagnostically after a positive screening test.

Selective screening (Risk factors) is the most widely used screening strategy for GDM (American Diabetes Association, 1998) (Metzger and Coustan, 1998). Risk factors may be found to be present during the first interview or visit at clinic. The main risk factors in screening for GDM include obesity, age greater than 25 years, family history of diabetes, history of previous GDM and ethnic group, depending on the criteria. Factors suggested in many guidelines include obesity, age, family history of DM and previous GDM. Various ethnic groupings have been recommended as high-risk for GDM in different guidelines. For instance, the current National Institute for Clinical Excellence (NICE) recommended only three ethnic groups, namely South Asian, Black Caribbean and Middle Eastern (National Health Service, 2008b). The Fourth International Workshop-Conference on Gestational Diabetes Mellitus and The Australasian Diabetes in Pregnancy Society suggested more high-risk ethnic groups, including Hispanic, African, Native American, South or East Asian , Pacific Islands and Indigenous Australian ancestry (Metzger and Coustan, 1998) (Hoffman et al., 1998). All common risk factors and various ethnic high risk groups are shown in Table 4.6. Kim and co-workers conducted a systematic review of risk factors associated with recurrence of GDM. The significant risk factor was ethnicity where the recurrence rate varied between 30% - 84%. Lower rates were found in non-Hispanic white groups, about 30% - 37%, and higher rates were found in minority groups, approximately 52% - 69% (Kim et al., 2007a).

In 2002, Scott and co-workers carried out a systematic review of risk factor screening tests for GDM. They found that the use of risk factor screening tests alone were associated with low sensitivity and specificity (Scott et al., 2002). In another systematic review, it was found that the use of risk factors alone as a screening test was associated with a low positive likelihood ratio of about 1.75 (women with risk factors

are only 1.75 times more likely to have GDM than those without). In other words, a diagnostic test only in women with positive risk factors will miss many women with GDM (Martine et al., 2007). For example, Marquette and colleagues reported about 50% sensitivity, 55% specificity, positive predictive values (PPV) of only 3% and negative predictive values (NPV) of risk factors screening (Marquette et al., 1985). In a study in North Carolina, Heltona and colleagues showed low sensitivity, specificity, PPV and NPV of about 69%, 68%, 5%, 55% respectively (Helton et al., 1997).

Table 4.6 Clinical risk factors for gestational diabetes mellitus

Clinical Risk Factors for Gestational diabetes mellitus
• body mass index (BMI) above 30 kg/m^2
• previous macrosomia baby weighing 4.5 kg or above
• previous gestational diabetes
• family history of diabetes (first-degree relative with diabetes)
• family origin with a high prevalence of diabetes:
– South Asian (specifically women whose country of family origin is India, Pakistan or Bangladesh)
– Black Caribbean
– Middle Eastern (specifically women whose country of family origin is Saudi Arabia, United Arab Emirates, Iraq, Jordan, Syria, Oman, Qatar, Kuwait, Lebanon or Egypt).
– South-east Asian
– Aborigine
– Hispanic
– African
– Pacific Islands and Indigenous Australian ancestry

On the other hand, approximately 40% – 60% of mothers with GDM have an invisible risk factor. The implication of this is that all pregnant women should be screened (ACOG, 1994). Another reason for screening all pregnant women is that women with GDM may exhibit no symptoms. In the systematic review, it was concluded that 50% of women with GDM would be indentified by using historic and clinical risk factors to detect those at risk from GDM (Scott et al., 2002). Weeks and colleagues stated that 43% of women with GDM remained undiagnosed after selective screening of patients using risk factors of obesity glycosuria, family history of DM, previous macorsomis and stillborn or anomalous foetus (Weeks et al., 1994). Poyhomen-Alho and co-workers reported 47% of cases were missed by selective screening when using 50g glucose challenge test (GCT) combined with risk factors (Poyhonen-Alho et al., 2005). Therefore, selective

screening should be used with another diagnostic test to confirm if disease is actually present.

One step and two step protocols

In general, the tests can be divided into screening tests and diagnostic tests. Screening and diagnostic tests are further divided into one-step and two-step. There has been debate about one-step and two-step protocols for the diagnosis of GDM (Ferrara and Kim, 2009). In the first case, in the one step strategy, a diagnostic test such as 50g GCT, 75g oral glucose tolerance test (OGTT) and 100g oral glucose tolerance test (OGTT) is performed without prior screening. In the second case, in the two step strategy, a diagnostic test is performed to confirm a positive screening test such as a random plasma glucose test, fasting plasma glucose test or glycosuria (American Diabetes Association, 2003) (Rey, 1999). For example, the one step approach by 50g GCT requires women to be tested in the morning after a 12 hour fast, following 3 days of a diet with at least 150g of carbohydrate per day. In the two step approach, only women with abnormal results in the initial screening are required to proceed to diagnostic screening, depending on the strategy.

4.9.2 Test techniques for the screening of gestational diabetes mellitus

Many test techniques for GDM have been described over the past decades. The screening and diagnostic test techniques have been developed to detect high levels of glucose or serum in defined circumstances. The test techniques for GDM can be divided into screening and diagnosis tests. Table 4.7 shows the list of test techniques that are used to identify GDM in pregnancy based on the guideline as discussed in previous section.

Table 4.7 Test techniques for the screening of gestational diabetes mellitus

	Test technique	Fasting : without food and drink (except for water)	Drink contains glucose (g)	Blood test drawn				Test		Suggestion
				0h	1h	2h	3h	screening	diagnostic	
1	Random Plasma Glucose Test (RPG)			√				√		
2	Fasting Plasma Glucose Test (FPG)	Overnight, for 10-16 hours before the test		√				√		
3	50g Glucose Challenge Test (50 GCT)		50		√			√	√	No food or drink during time collecting of blood sample
4	50g Oral Glucose Tolerance Test (50g OGCT)	Overnight, for 10-16 hours before the test	50		√			√	√	No food or drink during time collecting of blood sample
5	75g Oral Glucose Tolerance Test (75 OGTT)	Overnight, for 10-16 hours before the test	75	√	√	√			√	No food or drink during time collecting of blood sample
6	100g Oral Glucose Tolerance Test (100 OGTT)	Overnight, for 10-16 hours before the test	100	√	√	√	√		√	No food or drink during time collecting of blood sample

4.9.3 Screening test techniques for GDM

Screening tests for GDM include risk factors, random blood glucose (RBG), fasting blood glucose (FBG), and other tests. Abnormally high levels of glucose in a screening test may indicate GDM. However, a patient will probably be asked to undergo further diagnostic testing to confirm a diagnosis.

Random blood glucose (RBG)

Random blood glucose (RBG) is a simple, fast and inexpensive test for DM and GDM. The RBG is a measure of non-fasting glucose level. No glucose load need be given and the measurement is made randomly. Either random plasma glucose (RPG) or whole blood glucose (WBG) may be measured (Scott et al., 2002). In a national survey from the UK by Hanna and co-workers, it was shown that 25% of the respondents used a random glucose measurement to test for GDM (Hanna et al., 2008). Van and colleagues, in a systematic review of accuracy of the random glucose test as a screening tool for GDM, stated that random glucose measurement was inadequate as a screening tool for GDM based on limited studies. They reported of the accuracy of RBG, due to the small number of studies, when specificity approached 100%, sensitivity dropped to 20% – 30% (van Leeuwen et al., 2010).

Fasting Blood Glucose (FPG)

Fasting blood glucose (FPG) became popular for screening GDM after the greater acceptance of ADA and WHO guidelines for detecting DM. The FPG is a measure of blood glucose in a patient who has not eaten anything for at least 8 hours. The FPG test is most reliable when done in the morning. The risk of GDM is increased in patients who have positive FPG tests. The test is easy to perform and not very demanding for the subjects. In a study in a United Arab Emirates population, Agarwal and co-workers found the FPG using a threshold of ≥ 5.3 mmol/l had 48% sensitivity and 97.5% specificity and a threshold of equal to or lower than 4.3 mmol/l had 93% sensitivity and 38.5% specificity (Agarwal et al., 2000). Perucchini and co-workers concluded that the best threshold for FPG was 4.8 mmol/l, which give a sensitivity of 81% and specificity of 76% (Perucchini et al., 1999). The authors of the systematic review of screening for GDM stated that there is no conclusive data regarding the reproducibility of the sensitivity and specificity of the test. In addition, use of fasting glucose as a screening test is known to be unreliable for predicting a risk of macrosomia in a mother with either GDM or type 2 DM (Martine et al., 2007).

Other screening test techniques

However, other test techniques that are not so commonly used as screening tests for GDM are HbA1c, capillary blood glucose measurement with a hemocue, breakfast test, lunch test, glycosuria, blood fructosamine, and foetal abdominal circumference (Martine et al., 2007) In the past, convenient screening tests such as glycosuria, glycosylated hemoglobin A1c (HbA1c) and random plasma glucose (RPG) test were used to identify GDM. These were found to be not sensitive enough to measure the glucose level. Glycosylated hemoglobin showed low sensitivity to detect women with GDM and was not being recommended (Cousins et al., 1984). The breakfast and lunch tests have an advantage of using a standard test meal rather than an artificial glucose solution. This technique is not widely used worldwide and has limited research (Martine et al., 2007). The systematic review of screening for GDM and economic evaluation by Scott and co-workers stated that glycosuria is common in pregnancy unaffected by GDM. Glycosuria is routinely collected during pregnancy for other purposes than for detecting of GDM in pregnancy (Scott et al., 2002). One advantage is that screening for GDM as glycosuria is a cheap, convenient and standard nursing procedure (Hanna and Peters, 2002). Blood fructosamine showed very low sensitivity and seem to have little value as a screening test for GDM (Martine et al., 2007).

4.9.4 Diagnostic test techniques for GDM

The “gold standard” for diagnosis of GDM is 75g OGTT and 100g OGTT (Martine et al., 2007). The diagnostic test is more time-consuming and complicated than the screening test. This section focuses on 75g OGTT because the 100g OGTT is popular In North America, while the 75g OGTT version is used more in Europe and Asia. The 75g OGTT has been the international standard for the diagnosis of diabetes in the general population for several decades and has been endorsed by WHO in 1999 for use during pregnancy (WHO, 1999a) (Pettitt, 2001). The 75g OGTT is a measure of blood glucose after an overnight fast or fast for at least 8 to 12 hours (water only permitted). Before the three day period of the test, patients should maintain an adequate carbohydrate intake (150 - 200). Patients drink a glucose-containing beverage of 75g concentrate, finishing this within 5 minutes. This test involved 3 blood collections over 2 hours. Blood samples will be drawn at various intervals to measure glucose levels, while fasting, one and two hours after glucose intake. Table 4.8 provides a summary of the thresholds of 75g OGTT by different strategies commonly used to diagnose GDM.

Table 4.8 Gestational diabetes mellitus threshold on the 2 hour 75g OGTT test (mmol/l)

Interval time	EASD	WHO	CDA	ADIPS	SIGN	NICE	ICG	SIGN
Fasting	≥6	≥7	≥5.3	≥5.5	≥7	≥7	≥10	≥10
2 h	≥9	≥7.8	≥8.9	≥9.0	≥9	≥7.8	≥8.5	≥8.5

Organisation (WHO) World Health Organisation, (ADIPS) The Australasian Diabetes in Pregnancy Society,(ADA) American Diabetes Association, (CDA) Canadian Diabetes Association, (AGOG) American College of Obstetricians and Gynecologists ,(EASD) European Association for the Study of Diabetes, (SIGN) The Scottish Intercollegiate Guidelines Network,(NICE) the National Institute for Health and Clinical Excellence ,(ICG) International Consensus Guideline.

Different thresholds for the diagnosis of GDM are widely used, which can have an impact on the prevalence and case detection of GDM, as mentioned in section 3.2.3. No studies comparing the different cut-off points used in the 75g OGTT were found in this review, however studies are available that report the difference in cut-off points with the 100g GOTT and how they relate to changes in prevalence, as discussed below. Table 4.9 provides a summary of 100g OGTT and the strategies which are commonly used to diagnose GDM.

Table 4.9 Gestational diabetes mellitus threshold on the 3 hour 100g OGTT test (mmol/l)

Interval time	O'Sullivan-Mahan	NDDG	C&C	ADA	AGOG
Fasting	≥ 5.9	≥ 5.9	≥ 5.3	≥ 5.3	≥ 5.3
1 h	≥ 9.9	≥ 10.6	≥ 10.0	≥ 10.0	≥ 10.0
2 h	≥ 8.5	≥ 9.2	≥ 8.7	≥ 8.6	≥ 8.6
3 h	≥ 7.7	≥ 8.1	≥ 7.8	≥ 7.8	≥ 7.8

Organisation (C&C) Indicated Carpenter and Constant, (NDDG) National Diabetes Data Group, (ADA) American Diabetes Association, (CDA)Canadian Diabetes Association, (AGOG) American College of Obstetricians and Gynecologists

The prevalence of GDM was estimated in three studies using the C&C criteria and reported comparisons with NDDG criteria. A similar trend was outlined in the three studies such that prevalence's derived from C&C criteria were found to be greater than those from NDDG (Ferrara et al., 2002) (Karcaaltincaba et al., 2009) (Hadaegh et al., 2005). Schmidt and colleagues reported that 22 women (2.4%, 95% CI 2.0 – 2.9) were present with GDM by the ADA criteria alone and 260 women (7.2%, 95% CI 6.5 – 7.9) were diagnosed by WHO criteria (Schmidt et al., 2001). However, a study in Thailand, Soonthornpun and colleagues found that the 75g OGTT gave a significantly lower diagnostic yield of GDM than the 100g OGTT. Using the ADA criteria when the same threshold levels were measured, the results of prevalence were approximately 7.1% and 21.4% respectively (Soonthornpun et al., 2003).

The different screening and diagnostic thresholds also present with different outcome risks. In a Californian study, Ferrara and co-workers investigated whether woman who did not meet the NDDG criteria (Capenter and Coustan plasma Glucose thresholds) for GDM, but who exceed the ADA criteria are at risk of neonatal adverse outcomes. The results showed lower thresholds in ADA, in comparison to the NDDG criteria, which leads to a risk of complications, such as the risk of delivering an infant with macrosomia, hypoglycemia or hyperbilirubinemia (Ferrara et al., 2007). In 2003, Savona - Ventura and Chircop found a significantly increased risk of developing hypertensive disorders during pregnancy and caesarean section by using WHO criteria (Savona-Ventura and Chircop, 2003).

4.10 Current screening and diagnostic tests in Scotland

In Scotland, historically, there has been controversy over the best method of screening and specific diagnostic criteria for the detection and definition of GDM, with no accepted international consensus to this day. The Scottish Intercollegiate Guidelines Network produced guidelines in two different years, SIGN 2001 and SIGN 2010 (SIGN, 2001)

(SIGN, 2010). Although these two guidelines were published by the same organisation, they recommend different screening test methods and procedures.

Various screening test strategies have been proposed and discussed in section 4.7. Of these, four screening tests have been selected for further study, including SIGN 2001, SIGN 2010, NICE 2010 and Consensus 2010, each being developed by different organizations and employing different screening test techniques. Furthermore, the four screening test strategies include either a one or two step approach as well as either universal or selective screening procedures. Further details of this selection strategy are presented in chapter 7.

4.11 Conclusion

GDM is the most common medical complication found during pregnancy, with a large prevalence and imposing cost burden. GDM is defined as any degree of glucose intolerance with an onset or first recognition during the period of pregnancy. The high risk of maternal and neonatal adverse outcomes brought about by maternal hyperglycemia make it crucial to diagnose GDM during 24-28 weeks of gestation. Early diagnosis and appropriate treatment aim to control maternal glucose levels and may positively influence short and long term maternal and perinatal outcomes.

A variety of tests and thresholds for GDM have been proposed. Screening tests that are inadequate can result in missed cases of GDM and improper testing strategies may result in false positive results. Women may have less insulin resistance earlier in gestation and may show a false negative in screening, although screening at an inappropriate time may detect undiagnosed type 2 DM. This can lead to avoidable health care costs and cases of preventable maternal and neonatal morbidity.

Chapter 5 Prevalence of gestational diabetes mellitus: A systematic review by characteristic of screening tests

5.1 Introduction

Commensurate with the global increase in diabetes, generally the prevalence of GDM has increased in many countries associated with the rising prevalence of obesity and older age mothers. It is commonly perceived that the prevalence of GDM cannot be measured well by means of a systematic review. A problem with assessing the prevalence of GDM is that epidemiological data is captured with different test techniques. Numerous screening tests and guidelines for GDM have been developed and published over the last decade. A variety of different tests and guidelines are used to screen for GDM, and due to this the prevalence's for GDM that are reported in studies tend to vary considerably. In order to understand this variance, it is necessary to pay close attention to the characteristics of the screening tests and the population, such as test procedures, test guidelines and risk factors (e.g. ethnicity and age). Thus, the purpose of this chapter is to perform a systematic review to reassess the prevalence of gestational diabetes mellitus (GDM) considering not just screening test characteristics but also population characteristics. Systematic reviews are characterised by a clearly stated set of objectives and methodology, a systematic search that attempts to identify all articles to be included into the study, an assessment of the validity of the findings of the included studies, and a systematic presentation and synthesis of the characteristics and findings of the studies.

Accurate prevalence information required for health-care planning is important for several reasons as it provides essential knowledge in order to assess the burden of a condition within a population. Furthermore, accurate decision models require accurate estimates of prevalence and decision model outputs are highly sensitive to the prevalence. Additionally, in economic evaluations of screening tests prevalence is required as a probability in the construction of decision trees. Using prevalence and incident data it is possible to calculate expected numbers with and without disease for a given population. Hence, when modelling a health policy decision affecting a population, a decision tree that commences with branches for disease status can be used, enabling the direct application of conditional probabilities to the following branches, as mentioned in section 3.3.3.

The review begins with the search strategy as well as inclusion and exclusion criteria. The key words are detailed in this section. The next section looks at the quality

assessment of the prevalence studies included in this review. Following this there is a discussion of how the studies are selected and shown in the accompanying flowchart. The results of the review are discussed in terms of the various screening tests used in each study, the prevalence reports from the studies, and limitations and recommendations for future research. The final section offers a summary and conclusions.

5.2 Methodology

5.2.1 Search strategy

The two electronic databases, MEDLINE and EMBASE were used to search the literature on prevalence of GDM from January 1996 through July 2013. All reference lists from the main reports and relevant reviews were hand-searched. In addition, the following subject headings and text words, and their combinations were included in electronic database search strategy (incidence or epidemiolog* and prevalence and gestation* diabet* or GDM or pregnan* and diabet* and burden), as detailed in Appendix 1.

All retrieved studies were initially screened by title and abstract, then the exclusion/ inclusion criteria were applied. If it was concluded that an article should be included for systematic review, the full text of the article was obtained. Based on the full text manuscripts, studies were selected according to whether they matched predefined criteria.

5.2.2 Eligibility criteria

The newly published, Preferred Reporting Items for Systematic reviews and Meta-Analysed (PRISMA), provides guidelines for conducting systematic reviews, and suggests framing questions with five components to help facilitate the systematic review and meta-analysis process. Inclusion criteria are presented as five PICOS components namely, the patient population or the disease being addressed (P), the intervention or exposure (I), the comparator group (C), the outcome or endpoint (O), and the study design chosen (S) (Liberati et al., 2009). Predefined criteria were applied to choose the final list of papers to be included in the review based on the inclusion criteria presented as five PICOS components, as shown in Table 5.1.

Table 5.1 Eligibility criteria in PICOS components

PICOS components	Inclusion criteria
Populations(P)	Participants of any age of pregnancy. GDM was described as carbohydrate intolerance of varying degrees of severity, with onset or first recognition during pregnancy.
Intervention(I)	All screening test techniques and procedure
Comparative (C)	-
Outcome (O)	Papers which attempted to identity the prevalence of GDM as their primary objective were included, a subject has been identified for GDM papers provided a prevalent number (usually shown as a percentage) and 95% confident interval for the whole population or sub-populations of gestational diabetes mellitus were included
Study design chosen (S)	Cross-sectional study design or cohort studies.

Exclusion criteria included papers not in the English language, case studies based on a single case and prevalence numbers where patients were already known to have diabetes mellitus before pregnancy. Studies that report only prevalence trends for GDM over a number of years were also omitted. Additionally, papers consisting of abstracts only were omitted. However the potential for selection bias is recognised particularly with regard to English language and abstract exclusion.

5.2.3 Data collection process

All information about the methodology and results from each study were extracted using a data extraction collection form. Abstracted information included study characteristics and study results as well as other items; the study's location and population, PICOS, study's design, screening test criteria and threshold, and prevalence.

5.2.4 Methodological quality assessment

No generally accepted criteria for the quality assessment of prevalence studies were available. This systematic review applied and adopted the quality assessment list of previous systematic reviews of prevalence (Bishop et al., 2010) (Prins et al., 2002). Regarding the methodological quality of those studies, two aspects of validity are important: external validity and internal validity. External validity relates to the applicability of study results to other populations, whereas internal validity implies accurate measurement apart from random error. This quality assessment covers all aspects of the prevalence article including items deemed essential for the transparent reporting of a systematic review of prevalence studies. A checklist was designed that consists of 15 items: six items on internal validity; six items on external validity and; three items on informativity, as shown in Appendix 3. The quality assessments were scored for each item as positive (Yes) or Negative (No) and were not weighted for importance. If an item was

yes, its score equals one, whereas if an item was no, its score equals zero. The scores of the 15 items were summed up and presented as a total score for each study.

5.2.5 Statistical analysis methods

Descriptive statistics were used to summarise the quality of papers and the characteristics of the studies in accordance with the guidelines in the PRISMA statement (Liberati et al., 2009). The prevalence and its confidence interval (CI) were plotted for each study in a forest plot-type chart. Prevalence's were then grouped according to the characteristics of the respective screening tests and population, as detailed in Table 5.2; screening test procedures, risk factor (ethnic group) and screening test guidelines.

The screening procedures groups were broken down into two groups, universal and selective screening. Selective screening initially screens all pregnant women with risk factors. The studies were then grouped according to the main risk factors detailed in the work by National Institute for Clinical Excellence (NICE) and The Fourth International Workshop-Conference on Gestational Diabetes Mellitus (National Health Service, 2008b) (Metzger and Coustan, 1998). The risk factors are obesity, age greater than 25 years, family history of diabetes, history of previous GDM and ethnic group. Ethnic background was divided into two groups related to a high and low risk for GDM, based on the guidelines. The difference in screening test guidelines used may play a major part in the different incidences of GDM as well as screening procedures. However, if appropriate results were obtained from individual studies it was planned to estimate the overall prevalence of GDM using meta-analysis. Many systematic reviews contain meta-analyses, but not all.

Table 5.2 The characteristics of screening tests for GDM

Characteristics	Details
Screening procedure	Universal screening: screen all pregnant women by measurement of plasma glucose between 24 and 28 weeks using different test techniques
	Selective screening: screen initially all pregnant women with risk factors. In the next stage, women who show positive risk factors undergo further screening
Ethnicity group	High risk groups of GDM: South Asian, Back Caribbean, Middle Eastern, South-East Asian, Aborigine, Hispanic, African and Pacific Islands.
	Low risk groups of GDM: Other nationalities
Screening test guideline	Based on the guidelines used in the study

5.3 Study selection

The primary search by MEDLINE and EMBASE produced 366 papers. Removing duplicates and English language limitations, a total of 155 papers remained. Figure 5.1 illustrates the process of literature identification and study selection. Thus, 188 articles were reviewed for eligibility. Of those, 67 papers were excluded after screening title and abstract. 88 abstracts were accepted for further reading. 74 other studies were excluded for the following reasons: study population (n=9) outcome measurement (n=47), and study type (n=18). There were many reasons to exclude papers such as study population (included women with pre-existing DM and GDM or type 1 and type 2 diabetes), outcome measurement (the prevalence in postpartum diabetes and data not shown properly) and study type (report, review, and litter to edit paper). Only two articles were included following a hand search. Finally, 16 studies remained for final review. A PRISMA flowchart illustrates the process of study inclusion.

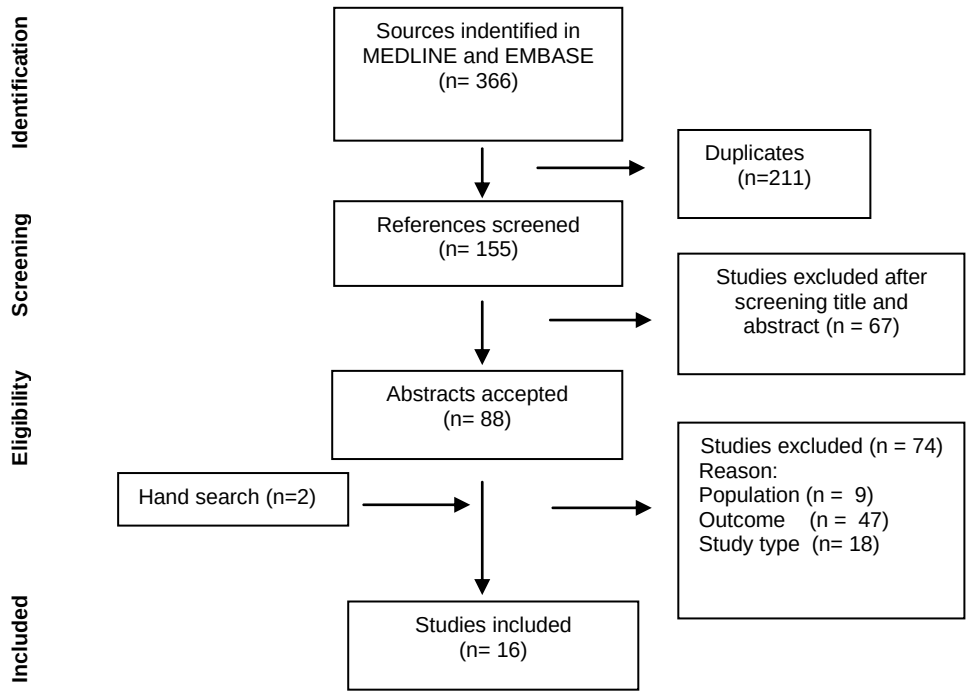


Figure 5.1 PRISMA flowchart of article selection

5.4 Methodological quality assessment

Table 5.3 displays the results of the quality assessment. On external validity, the average score was 4.6 (range 3-6), on internal validity the average score was 4.1 (range 3-6), and on informativity the average score was 2.4 (range 1-3). The quality assessment showed slightly high scores in all of the selected papers with more than half presenting with scores above 10. However some of the papers had lowers scores due to the effect of specific criteria. For example, external validity, the checklist items with the lowest scores were due to a lack of important population characteristics being specified and a response rate >70%. Regarding internal validity, the criterion with the lowest score was data collection, and thus was the most frequent reason for downgrading. Most of the studies used secondary data such as from medical records or databases. The last aspect of the quality assessment is informativity, and the item that scored the least was due to data collection not being properly described. It is recognised however that as scores are not weighted there may be an element of unreliability in the results.

Table 5.3 Year of published and quality assessment of selection studies

Reference		External validity							Internal validity							Informatively				
Number	Year	a	b	c	d	e	f	Sum	g	h	I	j	k	l	sum	m	n	o	Sum	
(Yue et al., 1996)	1996	+	+	-	-	-	+	3	-	+	+	+	-	-	3	+	+	-	2	8
(Godwin et al., 1999)	1999	+	+	+	+	+	+	6	-	+	+	+	+	-	4	+	+	+	3	13
(Rodrigues et al., 1999)	1999	+	+	+	+	+	+	6	-	+	+	+	-	-	3	+	-	+	2	11
(Seyoum et al., 1999)	1999	+	+	+	+	-	-	4	-	+	+	+	-	+	4	-	-	+	1	9
(Schmidt et al., 2000)	2000	+	+	+	+	+	-	5	-	+	+	+	-	+	4	-	+	+	2	11
(Ferrara et al., 2002)	2002	+	+	+	+	+	+	6	-	+	+	+	+	-	4	+	+	+	3	13
(Jimenez-Moleon et al., 2002)	2002	+	+	+	+	+	-	5	-	+	+	+	-	+	4	+	+	+	3	12
(Chanprapaph and Sutjarit, 2004)	2004	+	+	-	-	+	+	4	+	+	+	+	-	+	5	+	+	+	3	12
(Hadaegh et al., 2005)	2005	+	+	+	+	+	-	5	+	+	+	+	+	-	5	+	+	+	3	13
(Janghorbani et al., 2006)	2006	+	+	-	+	+	+	5	+	+	+	+	+	+	6	-	+	+	2	13
(Hossein-Nezhad et al., 2007)	2007	+	+	+	-	-	+	4	-	+	+	+	+	-	4	-	+	+	2	10
(Karcaaltincaba et al., 2009)	2009	+	+	+	-	+	-	4	-	+	+	+	-	+	4	+	+	+	3	11
(Pedersen et al., 2010)	2010	+	-	+	+	+	+	5	-	+	+	+	-	+	4	+	+	+	3	12
(Karcaaltincaba et al., 2011)	2011	+	+	-	-	+	-	3	-	+	+	+	-	-	3	+	-	+	2	8
(Moses et al., 2011)	2011	+	-	+	+	+	+	5	+	+	+	+	-	-	5	+	+	+	3	12
(Anzaku and Musa, 2013)	2012	+	-	+	-	+	-	4	-	+	+	+	-	+	4	+	+	+	3	10

Items a-o refer to table 1

5.5 Description of selected study populations

Descriptions of the populations included in the selected studies are given in Table 5.5. Five studies were conducted in general prenatal care clinic. Nine articles used secondary data including medical charts, computerised hospitalisation recorded and birth registration. Of the 16 studies, three were conducted in North America (Godwin et al., 1999) (Rodrigues et al., 1999) (Ferrara et al., 2002) and five Studies were conducted in Europe (Jimenez-Moleon et al., 2002) (Janghorbani et al., 2006) (Karcaaltincaba et al., 2009) (Pedersen et al., 2010). The remaining eight studies were conducted in Australia

(Yue et al., 1996) (Moses et al., 2011), Ethiopia (Seyoum et al., 1999), Thailand (Chanprapaph and Sutjarit, 2004), Brazil (Schmidt et al., 2000), Iran (Hadaegh et al., 2005) (Hossein-Nezhad et al., 2007) and Nigeria (Anzaku and Musa, 2013). Clinical prevalence was measured during 20 – 28 weeks of gestation. There was only one paper that reported the age range of subjects; teenage pregnancies age ≤ 19 in Ankara Turkey were studied by Karcaaltincaba and colleagues (Karcaaltincaba et al., 2011).

5.5.1 Diagnostic criteria for case inclusion

The criteria and test methods of screening for GDM are shown in Table 5.4. First, National Diabetes Data Group (NDDG) criteria were used in four studies (Rodrigues et al., 1999) (Ferrara et al., 2002) (Hadaegh et al., 2005) (Karcaaltincaba et al., 2009). The NDDG criteria screen by 1 hour, 50g GCT with cut-off point of ≥ 7.8 mmol/l and 3 hour 100g oral glucose tolerance test (OGTT); fasting, ≥5.8 mmol/l; 1 hour, ≥10.5 mmol/l ; 2 hour, ≥9.2 mmol/l; 3 hour, ≥8.1 mmol/l. Two studies from Iran and Turkey screened pregnant women by Carpenter and Coustan (C&C) criteria which proposed all women were screened for GDM by a 1 hour, 50g GCT, with cut-off point of ≥ 7.8 mmol/l (Hossein-Nezhad et al., 2007) (Karcaaltincaba et al., 2011). All women with positive screening test results underwent a 3 hour OGTT with 100g of glucose, with cutoff points of ≥ 5.8, ≥10.0, ≥ 8.6 and ≥7.8 mmol/l, respectively. A study in Canada used the screening strategy by The Third international work shop conference on GDM (Godwin et al., 1999). Four studies used World Health Organisation (WHO) guidelines to diagnose women with GDM (Seyoum et al., 1999) (Schmidt et al., 2000) (Janghorbani et al., 2006) (Pedersen et al., 2010) (Anzaku and Musa, 2013). WHO has recommended that all pregnant women have fasting plasma glucose tests exceeding ≥ 7.8 mmol/l, then followed by 2-hour 75g glucose solution tests with cutoff point; fasting ≥ 7.0 mmol/l; 2 hour, ≥ 7.8 mmol/l. Next criteria, from the Fourth international workshop conference on GDM, were used to screen patient with GDM in Granada, Spain (Jimenez-Moleon et al., 2002). This criterion has a cutoff point and screening test technique, which is from The Third International Work Shop Conference on GDM. In Australia a study used the criteria of The Australian Diabetes in Pregnancy Society (ADIPS) (Yue et al., 1996) (Moses et al., 2011). ADIPS was performed with 50g GCT. If the 1 hour blood glucose level exceeded ≥ 7.7 mmol/l, then a diagnostic 2 hour 75g OGTT was performed with ≥ 8.0 mmol/l. Lastly, a study in Thailand adopted the criteria of The Expert Committee on The Diagnosis and Classification of Diabetes mellitus to screen for GDM (Chanprapaph and Sutjarit, 2004). All women screened with 50g GCT and patients with result ≥ 7.8 mmol/l underwent 3 hour 100-g OGTT, with cutoff points of ≥ 5.3, ≥10.0, ≥ 8.6 and ≥7.8 mmol/l, respectively. Screening tests in this study can be applied either selectively or through universal screening. The Universal screening strategy is to screen all pregnant women by measuring their plasma glucose between 24

and 28 weeks by different test techniques. Conversely, selective screening strategy screens begin with all pregnant women by risk factors. Thus, women who have positive risk factors undergo further screening tests.

Table 5.4 Diagnostic test criteria and test methods of screening for GDM

Paper	Strategy screening	Year of published guideline	Test methods	Threshold values (mmol/l)				Screening procedure
				0 h	1h	2h	3h	
(Rodrigues et al., 1999, Ferrara et al., 2002), (Hadaegh et al., 2005, Karcaaltincaba et al., 2009),	NDDG	1979	50g GCT		≥7.8			
			100gOGTT	≥5.8	≥10.5	≥9.2	≥8.1	
(Hossein-Nezhad et al., 2007),(Karcaaltincaba et al., 2011)	C&C	1982	50g GCT		≥7.8			Selective
			100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8	
(Godwin et al., 1999)	3th	1991	50g GCT		≥ 7.8			Universal
			100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8	
(Seyoum et al., 1999),(Schmidt et al., 2000) (Janghorbani et al., 2006),(Pedersen et al., 2010), (Anzaku and Musa, 2013)	WHO	1998	50g GCT,		≥7.8			Universal
			75g OGTT	≥7.0		≥7.8		
(Jimenez-Moleon et al., 2002)	4th	1998	50g GCT		≥ 7.8			Selective
			100gOGTT	≥5.3	≥10.0	≥8.6	≥7.8	
(Yue et al., 1996) (Moses et al., 2011)	ADIPS	1998	50g GCT		≥7.8			Universal
			75g OGTT	≥5.5		≥8.0 ≥9.0		
(Chanprapaph and Sutjarit, 2004)	ECDC	2002	50g GCT		≥ 7.8			Universal
			100gOGTT	≥5.3	≥10.0	≥8.6	≥7.8	

Organisation (C&C) Indicated Carpenter and Constant, (NDDG) National Diabetes Data Group, (WHO) World Health Organisation, (ADIPS) The Australasian Diabetes in Pregnancy Society, (3rd)Third International work shop Conference on GDM, (4th) Forth International Workshop Conference on GDM, (ECDC) The Expert Committee on the Diagnosis and Classification of DM in USA

5.5.2 Prevalence of gestational diabetes mellitus

Studies were published between January 1966 and July 2013 and the prevalence estimates for GDM in the systematic review reveal that the pregnancies of 76,312 women ranged from 1.35% to 12.80%. However, the present analysis identified considerable heterogeneity in the prevalence of GDM and so refrained from calculating pooled estimates. Therefore, the individual articles are discussed.

There are slightly different populations in 16 studies. When comparing screening protocols, universal screening shows lower prevalence than selective screening. Four studies used selective screening which shows a prevalence range from 4.7% to 12.8% (Godwin et al., 1999) (Rodrigues et al., 1999) (Chanprapaph and Sutjarit, 2004) (Pedersen et al., 2010), as shown in Figure 5.2.

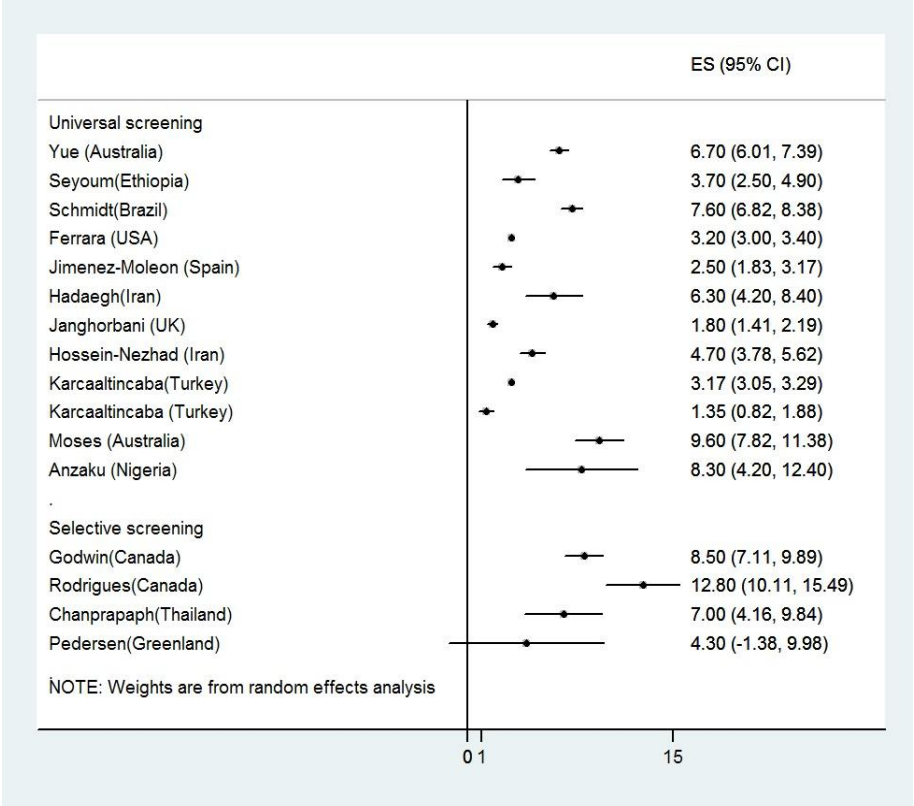


Figure 5.2 Prevalence of GDM groups by screening protocol characteristic

Ethnicity is a particularly important factor determining the prevalence of GDM. Figure 5.3 shows the distribution of GDM prevalence according to the different ethnic groups with both high and low risk factors for GDM. Ethnic groups with a high risk factor include South Asian, Back Caribbean, Middle Eastern, South-East Asia, Aborigine, Hispanic, African and Pacific Islands, while ethnic groups with a low risk factor include all other nationalities, as mentioned in Table 5.2. Prevalence of GDM when ethnicity is a low risk factor presented from 1.35% to 9.6 %. Whereas, the prevalence when ethnicity is a high risk factor of GDM varied between 4.7% and 12.8 %. Two of the studies conducted in Canada that determined the prevalence of GDM in Cree people, a Native American ethnic group, varied between 8.5% and 12.8 % (Godwin et al., 1999) (Rodrigues et al., 1999).

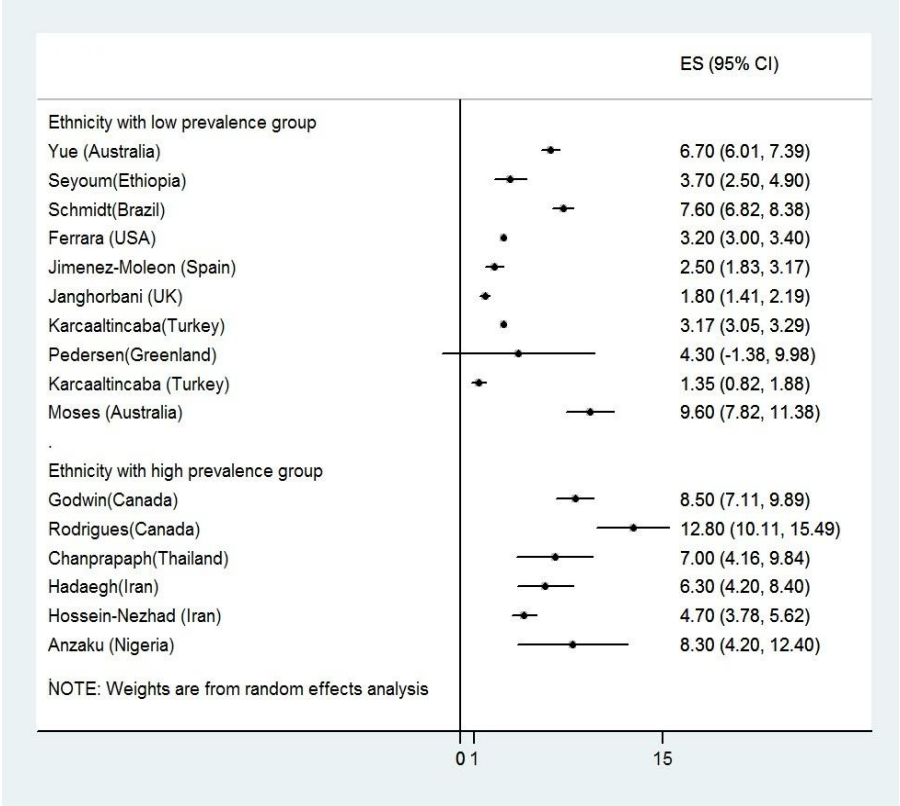


Figure 5.3 Prevalence of GDM group by ethnicity characteristic

There are four main screening guidelines in the review, namely Carpenter and Constant (C&C), World Health Organisation (WHO), National Diabetes Data Group (NDDG) as well as other guidelines. Prevalence varies widely amongst the screening guidelines and prevalence showed disassociation in the different screening strategies, as showed in Figure 5.4. Three studies report both NDDG and C&C strategies (Ferrara et al., 2002) (Hadaegh et al., 2005) (Karcaaltincaba et al., 2009). The prevalence of GDM for NDDG in the 3 studies was selected for analyses as the prevalence from the C&C guideline ranged from 4.8% to 8.9%, which is slightly higher than the NDDG guideline which ranged from 3.2% to 6.3%. Moreover, only one of the 16 studies presented two estimates of GDM prevalence: observed prevalence (number of confirmed cases among the total study) and expected prevalence (number of GDM cases among those women who had not been screened) (Jimenez-Moleon et al., 2002).

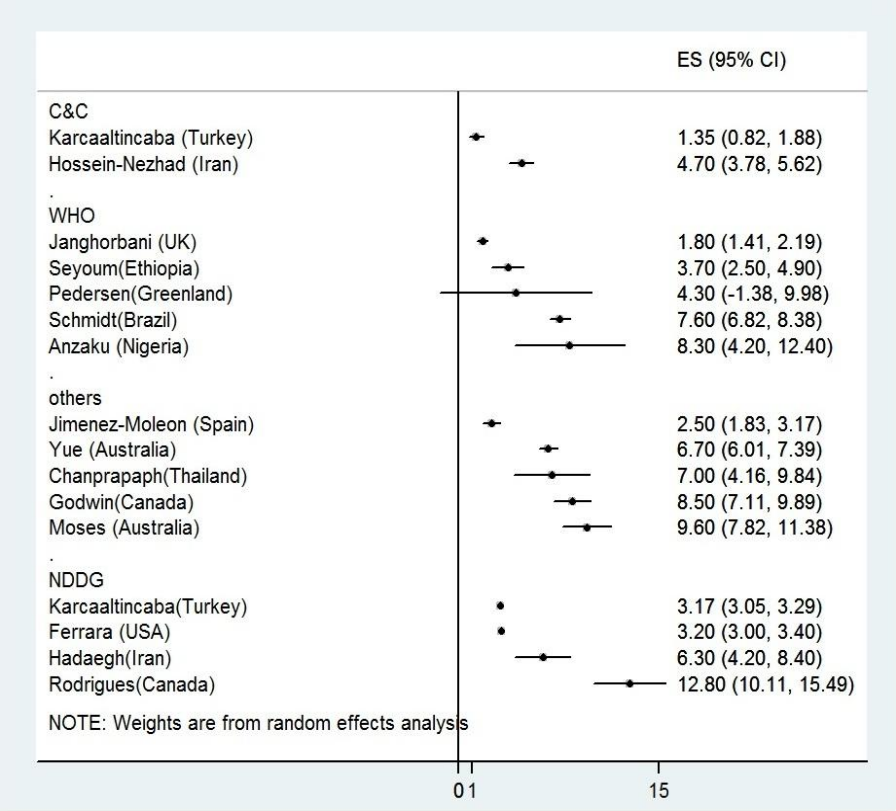


Figure 5.4 Prevalence of GDM group by guideline of screening test characteristic

5.6 Discussion

This is a systematic review of the literature focussing on the prevalence of GDM. Previously, available data in this field of research was summarised unsystematically or without a focus on a population based study and characteristic of screening test. There were 16 population-based studies of prevalence of GDM identified. Prevalence estimates for GDM depending on screening protocol and population ranged from 1.35% to 12.80%.

5.6.1 Selection of studies and data extraction

The prevalence rate reported for GDM is based on population-based studies that screen all pregnant women without GDM. Studies that reported trends to predict prevalence and those that reported prevalence in order to measure the accuracy of tests were omitted from this review. As one of the inclusion criterion of the systematic review was a requirement of a 95% confidence interval to estimate overall prevalence, the number of papers identified was limited to only 16. Two studies were found by a hand search out with MEDLINE and EMBASE.

5.6.2 Methodological quality assessment

This paper applied and adopted the quality assessment lists from previous systematic reviews of prevalence (Bishop et al., 2010, Prins et al., 2002), which are based on theoretical considerations as well as common sense, and can also be used for a systematic review of other conditions in the general population. These lists cover all the methodology required to construct the prevalence study. The distinction made between valid and invalid in the assessment is based on overall scores, and the use of cut-off points is arbitrary. It should be recognized; however, that some of the selection studies have a high number of negative scores (No), with a score equal to zero, as shown in Table 5.3. Additionally, the overall qualities of the studies were mixed, with more than half of the studies scoring > 50 %. One of the quality assessment lists had separate validity criteria in addition to the overall quality assessment, such as the representativeness of the study population (item d); 5 studies reported response rates lower than 70% and insufficient data were available on the representativeness of the population. However, quality seemed to be unrelated to reported prevalence. The method of study might not relate to prevalence rates.

5.6.3 Comparison of prevalence rates

This review shows that reports of prevalence for GDM vary considerably, as seen in the other systematic reviews outlined below, and that there are major population and screening test differences between the studies. It is unclear, however, whether this variance in prevalence reflects a true difference between populations, screening test procedure and screening test guideline. The differences of those characteristic precluded the comparison of prevalence rates in most of the studies. Previously, in 1998, King reported that the prevalence of women with GDM was between 1% and 16%, depending on diagnostic criteria and population studies. King's study used standardised population-based information of diabetes in adult communities worldwide produced by WHO (King, 1998). Another multi-stage cross-sectional prevalence study reported a range of 1% to 14% for GDM in women whose ages were between 15 and 49 years, in the USA in 1988 (Engelgau et al., 1995). Many factors have caused these two papers to produce different results for prevalence, such as screening test use and population group. Another problem regarding reports of prevalence of GDM is that epidemiology data is obtained from different test techniques. Numerous screening test guidelines for GDM have been developed and published within the last decade. Most countries have their own diabetes associations, each publishing their own guidelines for GDM, which may differ slightly.

However, the screening test strategies for GDM are still debated, and no consensus has been established yet. (Metzger and Coustan, 1998)

Differences in the prevalence of GDM may be related to variations in ethnicity. Hispanic, Native American, South or East Asian, Middle Eastern, South Asian and Black Caribbean women are in ethnic groups with relatively high rates of carbohydrate intolerance during pregnancy and of diabetes later in life (Metzger and Coustan, 1998) (Hoffman, 1998) (Bardenheier et al., 2013) and so there ethnic groups are also classified as high risk groups for GDM. Ethnic origin was the dominant influence on the prevalence of GDM in this review. A Canadian study undertaken among Cree women in the eastern James Bay region of northern Quebec, reported the highest prevalence in the reviewed papers of 12.8%, using the third international workshop conference on GDM criteria (Rodrigues et al., 1999). The authors also mention that this prevalence was at least twice as high as that reported in the general North American population. Another study in this review conducted on the same population, also reported a slightly high prevalence, about 8.5% (Godwin et al., 1999). Two additional studies also stated that the Cree ethnic group are the second highest prevalence of GDM reported in an aboriginal group worldwide (Sermer et al., 1995) (Magee et al., 1993). Similarly, In the United Arab Emirates a study of high risk populations reported a prevalence rate of about 19.3% (Agarwal et al., 2006).

Prevalence of GDM also varies between racial and ethnic groups within the same country. Two studies in Iran included in this review presented slightly different prevalence's, 6.3% and 4.3% (Hadaegh et al., 2005) (Hossein-Nezhad et al., 2007). The lower prevalence, reported in the study of rural populations, assumed a significantly different lifestyle (Hossein-Nezhad et al., 2007). Moreover, prevalence significantly differs for size and diversity of the population available for study in terms of geographic diversity (states, regions, and countries) and the definition used to identify women with GDM. Two studies in Australia reported a slightly high prevalence of 6.7% and 9.6% (Yue et al., 1996) (Moses et al., 2011). In the 2011 study, carried out in New South Wales City, it was not possible to record the women's country of birth at the time of collection. However the authors refer to the Australian Institute of Health and Welfare which reports that 83.5% of women giving birth in 2009 were born in Australia or were from countries with a predominately Caucasian background (Moses et al., 2011).

The systematic review of prevalence of GDM showed no association with screening strategy. Six popular screening strategies were used world-wide, including, in order of year proposed, O'Sullivan and Mahan (O'Sullivan and Mahan, 1964), NDDG (National Diabetes Data Group, 1979), C&C (Carpenter and Coustan, 1982), WHO (WHO, 1999b), the third international workshop conference on GDM (Metzger, 1991) and the

fourth international workshop conference on GDM (Metzger and Coustan, 1998). Screening and diagnostic test criteria have various associated screening test guidelines, which have been developed and published for GDM. Most countries have their own diabetes associations, these societies often publish guidelines for GDM, which may differ slightly. Therefore, the screening test strategies for GDM are still debated (Metzger and Coustan, 1998). Three studies that estimate the prevalence of GDM using the C&C criteria reported comparisons with the NDDG criteria. These three studies reported a similar trend, whereby prevalence derived from the C&C criteria were reported to be higher than the prevalence from NDDG (Ferrara et al., 2002) (Karcaaltincaba et al., 2009) (Hadaegh et al., 2005). The average increase in the prevalence of GDM, reported in two of the three studies, was 50% (Ferrara et al., 2002) (Ricart et al., 2005). NDDG suggested a two step screening procedure in 1979; Women are screened by 50g GCT with “hour 1” oral glucose tolerance test and undergo a diagnostic test “hour 3”, 100g OGTT, after abnormal screening tests. On the other hand, In 1982, The C&C proposed two step screening; “hour 1”, 50 g GCT with cutoff point ≥7.8, and then diagnoses by “hour 3”, 100g OGTT; fasting, ≥5.3 mmol/l; “hour 1”, ≥10.0 mmol/l; “hour 2”, ≥8.6 mmol/l; “hour 3”, ≥7.8 mmol/l. In both sets of criteria, GDM was diagnosed by at least two plasma glucose measurements exceeding the reported cut off point during the diagnostic tests. In addition, both NDDG and C&C strategies revised the O’Sullivan and Mahan criteria, converting whole blood values to plasma values. The diagnostic criteria from NDDG have been used most often, but some maternal care clinics apply C&C criteria, which set the threshold for normal at a lower value (Turok et al., 2003) (National Diabetes Data Group, 1979). Prevalence of GDM was therefore higher when using C&C criteria compared with NDDG criteria.

The distribution of classical risk factors in the general population of pregnant women is an important consideration in determining the optimal screening protocol. Selective screening screens all pregnant women with more than one risk factor, showing higher prevalence than universal screening. In this review, it was shown that papers conducted on high risk groups presented significantly higher prevalence rates (Godwin et al., 1999) (Rodrigues et al., 1999) (Chanprapaph and Sutjarit, 2004) (Pedersen et al., 2010). The results of selective screening tests including all mothers with high risk factors (as confirmed by diagnostic tests) demonstrated higher prevalence rates than universal screening tests. This is evidenced in the selective screening tests performed on high risk ethnic groups in two Canadian studies (Godwin et al., 1999) (Rodrigues et al., 1999). Consequently, the higher prevalence rates from selective screening seen in this review may have been driven by ethnicity.

Additionally, three retrieved studies reported subject ages (Schmidt et al., 2000) (Ferrara et al., 2002) (Karcaaltincaba et al., 2011). There was little evidence to conclude that GDM prevalence estimates are dependent on the age of pregnant women studied in those three studies. However, pregnant mothers over 35 years old are classified as a high-risk group for GDM. One of those presented the lowest prevalence as 1.3%, which was found in teenage pregnancies in Ankara, Turkey, for women who were aged less than 19 years. Similarly, a study in Northern California reported a lower prevalence in pregnancy age < 25, of about 1% with NDDG criteria and about 1.7% based on the C&C criteria (Ferrara et al., 2002). The prevalence of GDM increases with the increased age of pregnant women (Getahun et al., 2008). Thus, age difference of screened mothers was shown to result in various prevalence estimates of GDM.

Variation in the prevalence of GDM can be explained by screening tests and population characteristics. Reports of prevalence among different ethnic groups in the Scottish population are not available in any known studies. In the UK, NICE reports the estimated incidence of gestational diabetes as 3.5% based on assuming age-standardised prevalence of 3.5% with type 2 DM and on a population of 39,57,157 aged 18 years or older (NICE, 2008a). Dornhorst and colleagues reported that women from ethnic groups other than white had a higher frequency of gestational diabetes than white women (2.9% vs 0.4%, $p < 0.001$) in the UK. Compared to white women, the relative risk of gestational diabetes in the other ethnic groups in the UK, was: Black 3.1 %, South East Asian 7.6 %, Indian 11.3 %, and miscellaneous 5.9 % (Dornhorst et al., 1992). This study confirms that different ethnic groups in the UK have dissimilar susceptibility with regard to the development of GDM. This result clearly shows that women in ethnic minority groups have an increased risk of GDM in UK. The largest ethnic population group in Scotland, at around 80% of the population, is white, and is similar to England & Wales (The Scottish Government, 2011) (Office for National Statistics, 2012).Therefore, prevalence of GDM in Scotland is taken to be approximately 3.5%.

5.6.4Limitation and recommendation for future research

Publication bias was not investigated in this review, as funnel plots were not considered appropriate due to the variation across studies. It is unlikely that the set of published papers are biased with respect to the prevalence reported. However, it is possible that some studies were not identified in the searches if they were not published in mainstream journals. However, studies brought together in a systematic review will differ (Glasziou and Sanders, 2002). In this systematic review it was not possible to pool prevalence estimates of GDM because of the considerable heterogeneity across individual studies. There are many factors which impact the heterogeneity of GDM. Those

factors may include socio-demographic features of the population such as: age, the ethnic group and screening test procedure. Another factor that may have had an impact is the number of papers identified, which was limited to only 16. This was because one of the inclusion criteria of the systematic review was the requirement of a 95% confidence interval to estimate overall prevalence, and many papers are poor at reporting prevalence within a 95% confidence interval.

Variability in ethnicity of those screened for GDM resulted in wide differences in prevalence. Even more concerning is that the diversity in screening test criteria for GDM usually results in varying prevalence. Moreover, unclear risk factors for GDM in each study such as family history of diabetes mellitus before pregnancy, history of macrosomia in previous pregnancy, history of preeclampsia during gestation and weight gain during pregnancy may have resulted in heterogeneity in this study. A move towards the application of similar strategies for screening tests for similar populations may reduce the heterogeneity of prevalence for GDM.

5.7 Conclusion

This review has shown some of the influences on variation of prevalence for GDM. The present review indentified several studies. These studies were heterogeneous in methodological quality and results. Variation in over half of the studies can be explained by ethnicity and screening test strategies. Other important factors were whether universal or selective screening protocols for GDM were adopted and whether 75g or 100g OGTT were used. The impact of these identified factors on prevalence estimates should be further investigated as they may be acting as proxies for other influences on prevalence. For example, although the literature on prevalence and the incidence of GDM is varied, there should be a general consensus for GDM screening strategies with regard to population. Accurate prevalence for GDM is as important as health-care planning and epidemiological research, as they provide essential knowledge to assess the burden of a condition within a population. Documentation of GDM‘s impact on quality of life and costs help to inform public health planning. Moreover, it is essential to begin base line prevalence rates, so that researchers can monitor trends.

Table 5.5 Description of the characteristics of the study

Author	Year	Country	Study design	Source of population	Ethnicity	Age	Total (N)	Duration of Follow-up	Time of screening (weeks)	Strategy	Screening	Prevalence (95% CI)
(Yue et al., 1996)	1996	Australia	Retrospective study	GDM in Obstetrics clinics	Mix	NR	5243	2 years	24-28	ADIPS	Universal	6.7 (6.0-7.4)
(Godwin et al., 1999)	1999	Canada	Retrospective	Medical chart who give birth at hospital	Cree women	NR	1401(1298)	8 years	24 -28	3rd	Selective	8.5 (6.0-9.9)
(Rodrigues et al., 1999)	1999	Canada	Cross-sectional study	Patient charts on pregnancies	Cree women	NR	704(654)	1 year	22 and above	NDDG	Selective	12.8 (10.1-15.5)
(Seyoum et al., 1999)	1999	Ethiopia	Community base study	Mother who register a health institution	Ethiopia	NR	(930)890	NR	24 and above	WHO	Universal	3.7 (2.5 – 4.9)
(Schmidt et al., 2000)	2000	Brazil	Cohort study	Prenatal care clinics of National Health Service	Brazilian	≥ 20	5564(5004)	5 years	20 - 28	WHO	Universal	7.6 (6.9 – 8.4)
(Ferrara et al., 2002)	2002	USA	Cross-sectional study	Computerized hospitalization record	Mix	14-49	28330(26,481)	1 year	24-28	NDDG	Universal	3.2 (3.0-3.4)
(Jimenez-Moleon et al., 2002)	2002	Spain	Retrospective cohort study	The birth register of all the pregnancy	Spanish	NR	2780(2574)	1 years	24-28	4th	Universal	2.53 (1.9 – 3.2)
(Chanprapaph and Sutjarit, 2004)	2004	Thailand	Retrospective study	Medical record of pregnant women at the antenatal care	Thai	NR	1000	15 Months	24 -28	ECDC	Selective	7.05 (4.8 – 9.9)
(Hadaegh et al., 2005)	2005	Iran	Prospective study	GDM in Obstetrics clinics	Iran	NR	800(700)	2 years	24 -28	NDDG	Universal	6.3 (4.7 – 8.4)

NR = not reported, **Organisation** (C&C) Indicated Carpenter and Constant, (NDDG) National Diabetes Data Group, (WHO) World Health Organisation, (ADIPS) The Australasian Diabetes in Pregnancy Society, (3rd)Third International work shop Conference on GDM, (4th) Forth International Workshop Conference on GDM, (ECDC) The Expert Committee on the Diagnosis and Classification of DM in USA

Table 5.5 (Continue)

Author	Year	Country	Study design	Source of population	Ethnicity	Age	Total (N)	Duration of Follow-up	Time of screening (weeks)	Strategy	Screening	Prevalence (95% CI)
(Janghorbani et al., 2006)	2006	UK	Prospective study	A clinical based in Plymouth UK	English	NR	4,942	2 years	26-28	WHO	Universal	1.8 (1.4 – 2.2)
(Hossein-Nezhad et al., 2007)	2007	Iran	Cross-sectional study	All women referred to antenatal clinics	Iran	NR	2416	18 months	24 -28	C&C	Universal	4.7 (3.9 – 5.6)
(Karcaaltincaba et al., 2009)	2009	Turkey	Retrospective study	Hospital information system recorded	Turkish	NR	21531	2 years	24-28	NDDG	Universal	3.1 (3.0 – 3.2)
(Pedersen et al., 2010)	2010	Greenland	Retrospective study	Birth-log and medical recorded	Greenlander	NR	268(233)	1 year	NR	WHO	Selective	4.3 (0 – 10)
(Karcaaltincaba et al., 2011)	2011	Turkey	Retrospective study	Data women pregnancy ≤ 19 years	Turkish	≤ 19	1659	2 years	24-28	C&C	Universal	1.35 (0.7 – 1.8)
(Moses et al., 2011)	2011	Australia	Prospective study	GDM in Obstetrics clinics	Mix	NR	1,422(1,275)	1 year	24 -28	ADIPS	Universal	9.6(8.1-11.4)
(Anzaku and Musa, 2013)	2013	Nigeria	Cross-sectional study	GDM in Obstetrics clinics	Nigerian	NR	265	4 months	24 -28	WHO	Universal	8.3(5.2-12.4)

NR = not reported, **Organisation** (C&C) Indicated Carpenter and Constant, (NDDG) National Diabetes Data Group, (WHO) World Health Organisation, (ADIPS) The Australasian Diabetes in Pregnancy Society, (3rd)Third International work shop Conference on GDM, (4th) Forth International Workshop Conference on GDM, (ECDC) The Expert Committee on the Diagnosis and Classification of DM in USA

Chapter 6 Literature review of economic evaluation in GDM

6.1 Introduction

Screening tests for gestational diabetes mellitus (GDM) are widely reported to be a cost effective intervention. This implies that any additional benefits achieved by preventing adverse complications in both mother and offspring and its consequences are worth the additional cost to a health service implementing screening and treatment for GDM. In order to provide valid information to inform policy makers, the economic evaluation of screening tests for GDM must be based on an appropriate model of the disease process and appropriate estimates of the incidence of the disease and its consequences. Over the past decade, decision making in healthcare has increasingly been based on considerations of cost effectiveness. This trend is reflected in most national guidelines concerning GDM screening. Several cost and cost-effectiveness studies of screening tests for GDM have been published in recent years. In the absence of comprehensive reviews of costs and economic evaluations of screening tests for GDM, the aim therefore of this chapter is to present a review of published papers about the cost and cost-effectiveness of screening tests for GDM.

The review is split into two sections with the first section focusing on cost analysis. Cost analysis is a partial form of economic appraisal; it looks only at the costs of the programmes and does not provide information on the health outcomes of interest. However the attention that the cost analysis studies have received to date from analysis is relatively low. Therefore, cost analysis was separated from the economic evaluations (CEA, CUA) because cost analysis is not a full economic evaluation. However, the cost analyses do give important information on the key costs associated with screening tests for GDM. This section discusses the key concepts of the economic evaluations identified in chapter 2, including perspectives, patient groups, comparators, outcome measures, the results of the evaluations and the handling of uncertainty, all of which were used to assess the studies.

In the second section, economic evaluations of screening tests for GDM are reviewed. The fundamental task of any economic evaluation is to compare the costs and consequences of the alternatives being considered. However although economic evaluations consider costs in a similar way, their handling of benefits tend to differ. Therefore, in this review, the economic evaluations that have developed an economic model in terms of cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) are

examined. CEA is one form of full economic evaluation where both costs and consequences in terms of natural effects or physical units of health intervention are examined. However, the outcomes of CEA may be insufficient. For this reason, CUA is now becoming more popular, because it is recognized that most health interventions impact upon both the length and quality of life. In addition to the key concepts outlined above, a description of the model's structure, design and sources of input data and the assumptions used within each model are discussed.

6.2 Methodology

6.2.1 Search Strategy

A structured search strategy that incorporated relevant keywords to identify relevant articles in economic evaluations was based on the NHS Economic Evaluation Database (NHS EED) (NHS EED, 2007). The search keywords of screening tests for GDM were combined with the NHS EED search filter in order to find cost and cost-effectiveness analysis papers of screening tests for GDM, as detailed in Appendix 1. A systematic literature search was carried out for literature published between January 1996 and December 2013 using MEDLINE, EMBASED, NHS Economic Evaluation Database (NHS EED), The Health Technology Assessment Database and the Cochrane Database of Systematic Reviews. In addition, the reference lists from key papers were hand-searched as well as searches for studies and papers. All retrieved studies were screened based on their titles and abstracts. Exclusion/ inclusion criteria were then applied and only the papers satisfying these criteria were selected. Predefined criteria were used to choose the final list of papers to be included in the systematic review.

6.2.2 Eligibility criteria

The inclusion criteria are presented as five P.I.C.O.S. components as mentioned in the previous chapter (Liberati et al., 2009). The inclusion criteria applied to the papers are shown in Table 6.1. Articles fulfilling these criteria were included in the systematic review, while exclusion criteria included papers that were not written in English and those that only consisted of abstracts.

Table 6.1 Inclusion criteria in PICOS components

PICOS components	Inclusion criteria

Populations(P)	All pregnant women
Intervention(I)	Any form of screening test intervening for GDM, including both universal and selective screening
Comparative (C)	No testing or alternative testing strategies (in economic modelling),
Outcome (O)	Information about cost and effectiveness were reported. The relevant costs of GDM have to be reported, as well as their method of calculation. Whether or not the effectiveness of the study is properly described. Studies reporting the results of economic models should be included such as; incremental costs per case detection, incremental costs per case prevention, incremental costs per life year gained and incremental coats per QALY/ DALY gained / adverted
Study design chosen (S)	Economic evaluation incorporating cost-effectiveness analysis and cost-utility analysis, primary study of the costs

6.2.3 Data collection process

All information about the methodology and results from each study were extracted using a data extraction/collection form. Details relating to the characteristics and results of included articles were extracted, such as the study's location and population, P.I.C.O.S, details of intervention and comparators, study design, sources of resource use, unit costs and consequences data, time horizon for cost and consequence and results of sensitivity analysis undertaken. Each article was summarised according to the above lists.

6.2.4 Methodological quality assessment

Despite the growing use of decision analytic modeling in CEA, Sculpher and colleagues developed lists for assessing quality in decision models (Sculpher et al., 2000). In brief, the tool consists of 35 items, which are broken down by structure (disease states, options, time horizon, and cycle length), data (identification, incorporation, handling uncertainty) and consistency (internal and external). Philips and colleagues proposed guidelines for quality assessment in decision-analysis models that are based on Sculpher's checklist (Philips et al., 2004). The Philips checklist contains 23 items. This review used these reporting guidelines as the basis for quality assessment. In addition, the Philips checklist provides a structure for any critical appraisal of an economic model and is used to complement the Drummond checklist (Drummond and Jefferson, 1996). Though this review includes cost analysis which is a partial form of economic appraisal it only looks at the costs of the programs and does not provide information on the health outcomes of interest. Currently there is no accepted criterion for the quality assessment of cost analysis studies available. Therefore, this systematic review will ignore the critical appraisal of cost analysis papers.

In order to summarise and assess the findings of the quality assessment, a formal grading of the evidence was developed for this systematic review and presented as "++",

"+" or "-" based on the system developed for the National Institute for Clinical Excellence (NICE) Public Health Guidance (NICE, 2012). There were three possible answers to each of the 23 questions in the Philips checklist. These were "yes", "no" and "not applicable". A scoring system for the checklist has been devised for this review, and each study has an individual denominator: firstly, a score of one was given to the "yes" response to each of the 23 questions in the Philips checklist. Secondly, all questions resulting in a "not applicable" answer were removed from the denominator. For example, a study that resulted in "yes" responses to all the questions on the Philip checklist would get a score of 100% (23/23). Whereas a study that resulted in one "non applicable" response, one "no" response and "yes" responses to the rest of the checklist questions, the total score would be 95.38% (21/22) This means the total score gives the proportion of "yes" responses for each study assessment. Studies with a score of equal or greater than 80% (≥ 80) were presented as "++"; studies with a score of equal or greater than 50% (≥ 50%) and lower than (< 80%) were presented as "+" and studies with a score of lower than 50% (< 50%) were presented as "-".

6.2.5 Statistical analysis methods

Descriptive statistics were used to summarise the quality of papers and the characteristics of the studies in accordance with the guidelines in the PRISMA statement, as detailed in Appendix IV (Liberati et al., 2009). In the first instance, the quality of the economic evaluations of the studies was assessed. Papers that present as "++" and "+" were reviewed in full and marked with a pass. In the second stage the characteristics of the studies were recorded. This systematic review divided the articles for analysis into two sets; those with cost analysis studies and those with cost-effectiveness studies.

6.3 Study selection

A detailed diagram of the review process is presented in Figure 6.1.The electronic searches yielded 696 journal article references. Removing duplicates and English language limitations, a total of 651 papers remained. After a title and abstract review, 606 studies were excluded. From the remaining articles, 45 were accepted for future reading. 28 articles were omitted for the following reasons: outcomes did not perform a cost and cost-effectiveness analysis (n=24), no full text available (n=2) and study design (n=2). Finally, 17 articles were reviewed in decision analysis modeling.

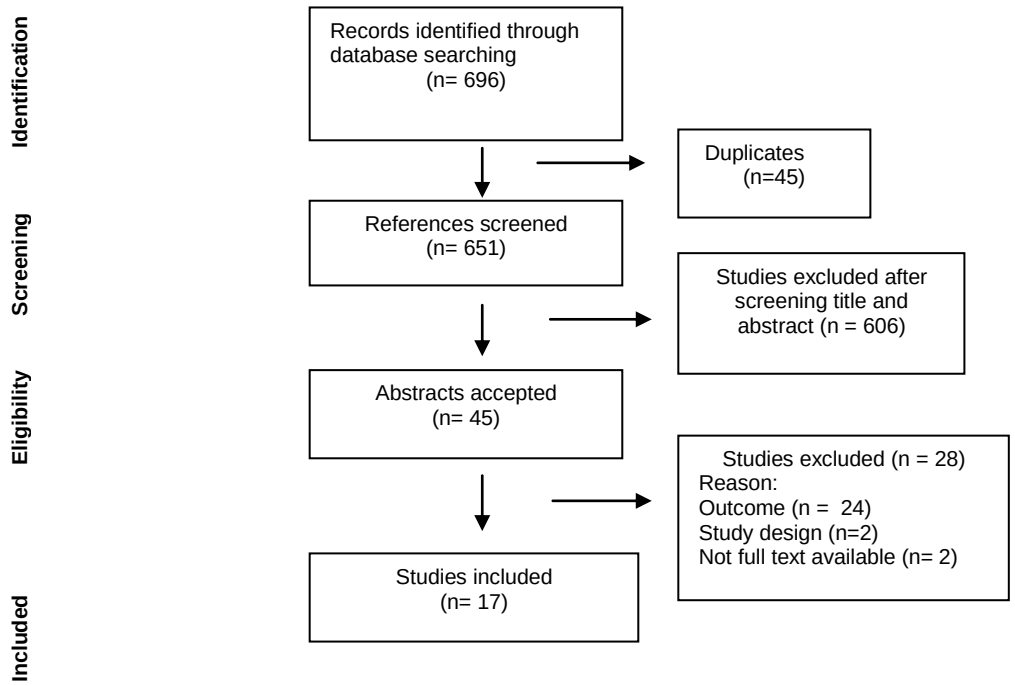


Figure 6.1 Flowchart of article selection

Table 6.2 below provides a concise summary of the features of each of the 17 economic evaluations, including the type of study (cost analysis, cost effectiveness analysis and cost utility analysis) the structure of the study (observational, RCT and model) and the main outcome measures used (case detection, accuracy of tests, case prevention and adverse complications, quality adjusted life year and disability adjusted life year). There has been a rapid increase in the number of cost analyses and economic evaluations published about GDM in recent years. Cost analysis was the most popular technique in the early years of GDM; approximately half of the evaluations were conducted using cost analysis (9/17). Of the later studies, most of the papers focus on modeling and the use of the QALY as an outcome measurement. Five papers were conducted using cost utility analysis (5/17). Seven used data from observational studies and eight used economic modeling. Case detection was the most frequently used outcome measurement, although others were also identified, including case prevention and adverse complications, QALY and DALY. As mentioned in the inclusion criteria, two types of economic evaluations which include cost-minimisation analysis and cost-benefit analysis are omitted from this systematic review. Therefore, two papers with these two types of economic evaluation were not taken into account (Kitzmiller et al., 1998) (Meltzer et al., 2010).

Table 6.2 Summary of GDM economic evaluations

Author	Year	Type			Structure				Outcome			
		CA	CEA	CUA	Obs	RTC	Sys	Model	CD	CP	QALY	DALY
Moses et al	1997											
Lemen et al	1998											
Poncent et al	2002											
DI Cianni et al	2002											
Scott et al	2002											
Larijani et al	2003											
Nicholson et al	2005											
Moss et al	2007											
NICE	2008											
Round et al	2011											
Lohse et al	2011											
Gillespie et al	2011											
Werner et al	2012											
Mission et al	2012											
Gillespie et al	2012											
Agarwal	2012											
Marseille	2013											

CA = Cost analysis, CEA = Cost effectiveness analysis, CUA = Cost utility analysis,
Obs = Observational, RCT = Randomize control trial, Sys = Systematic review, CD = Case detection,
CP = Case prevention and adverse complicaiton, QALY = Quality adjusted life year, DALY = Diability adjusted life year

6.4 Methodological quality assessment

For the economic evaluation studies, eight papers were assessed in total for their overall quality. Generally the studies performed well, with one study being graded "+", seven studies graded as "++" and no studies were graded "-". The results for each of the studies are presented in the summary table of evidence (Table 6.3). The last column in Table 6.6 displays the grading of the quality assessment.

Table 6.3 Author and quality assessment of selection studies

#	Dimension of quality	(Poncet et al., 2002)	(Nicholson et al., 2005)	(NICE, 2008b)	(Round et al., 2011)	(Werner et al., 2012)	(Lohse et al., 2011)	(Mission et al., 2012)	(Marseille et al., 2013)
	Structure								
1	Statement of decision problem	√	√	√	√	√	√	√	√
2	Statement of perspective	√	√	N/A	√	√	N/A	√	N/A
3	Rational for structure	√	√	√	√	√	√	√	√
4	Structure assumptions	√	√	√	√	√	√	√	√
5	Strategies/comparators	√	√	√	√	√	N/A	√	N/A
6	Model type	√	√	√	√	√	√	√	√
7	Time Horizon	N/A	N/A	N/A	√	√	N/A	N/A	N/A
8	Disease states	√	√	√	√	√	√	√	√
9	Cycle length	N/A	N/A	√	√	√	N/A	√	N/A
	Data								
10	Data identification	√	√	√	√	√	√	√	√
11	Data modeling	√	N/A	√	√	√	√	√	√
12	Baseline data	√	N/A	√	√	√	√	√	√
13	Treatment effects	√	√	√	√	√	√	√	√
14	Costs	√	X	√	√	X	√	√	√
15	Quality of life weights	X	√	√	√	√	N/A	√	√
16	Data incorporation	√	X	√	√	√	N/A	√	N/A
17	Assessment of uncertainty	√	√	√	√	√	√	√	√
18	Methodological for uncertainty	√	√	√	√	√	√	√	√
19	Structural uncertainty	√	√	√	N/A	N/A	√	√	√
20	Heterogeneity	X	X	√	X	X	X	X	X
21	Parameter	√	√	√	√	√	N/A	√	√
	Consistency								
22	Internal consistency	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
23	External consistency	N/A	√	√	√	√	√	√	√
	Total	(17/19)	(14/18)	(20/20)	(20/21)	(19/21)	(14/15)	(20/21)	(16/17)

√= " Yes", X = "No", N/A="Not applicable"

6.5 Review of the costs analysis studies

The studies were published between 1997 and 2012. Nine of the papers used non-modeling analysis, which is not a full economic evaluation and is discussed in this section. Seven papers conducted the cost analysis and one paper conducted the cost effectiveness analysis (Di Cianni et al., 2002). One study conducted a systematic review and an economic evaluation of GDM (Scott et al., 2002). A summary of each of the nine studies is discussed within this section, as shown in Table 6.4.

Screening test strategies

Various screening test methods were used in cost analyses to identify mothers with GDM, such as the 50g glucose challenge test (GCT), fasting plasma glucose (FPG), 2h 75g oral glucose tolerance test (OGTT) and 3h 100g OGTT. The most common standard test in all of the studies is 75g OGTT and 100g OGTT.

Methods

Population numbers in each of the studies varied substantially from 907 to 2416. The populations were not different across the nine studies. The common populations were women with singleton pregnancy that were without diabetes before pregnancy. Of these, the main risk factors included women with a family history of diabetes, family origin with a high prevalence of diabetes, previous gestational diabetes, previous children with macrosomia, glycosuria, maternal age of 35 years or older and obesity which was based on criteria.

No references to perspectives were made in the eight papers. Only one paper reported the perspective of the study which was the health care provider (Moss et al., 2007). However, of those papers, the type of costs and resources that were mainly used in the papers were either only direct costs or direct and indirect costs of medical care, such as drug costs, laboratory costs, staff administration costs and even allocated hospital overhead costs.

Of those papers, four papers contain evaluations in which two or more interventions are compared, but in which the costs and outcomes of each alternative are not examined simultaneously (Di Cianni et al., 2002) (Larijani et al., 2003) (Moss et al., 2007) (Agarwal et al., 2012). There are four papers that are conducted without comparators. All pregnant women are usually offered universal screening based on

specific criteria (Moses et al., 1997) (Lemen et al., 1998) (Kitzmiller et al., 1998) (Gillespie et al., 2011). In order to compare screening programmes for GDM in cost analyses, screening tests were classified into two or three groups according to the objective of the study. Selective screening was the most popular comparator used in the three studies. However, one study of cost and consequence compared pregnant women in two groups; intervention and routine (Moss et al., 2007).

Two outcome measures were used within the studies. The most frequently used was case detection seen in nine non-modeling studies. In addition to this, others included cost per additional serious perinatal complication prevented and per perinatal death prevented (Moses, 2010). Six studies used data from observational studies, two used data from randomised clinical trials and one used literature review.

One-way sensitivity analysis was used in two studies (Larijani et al., 2003) (Gillespie et al., 2011). This analysis showed that an increasing prevalence of GDM led to increases in total cost estimates. Two additional factors which can change cost estimates are the sensitivity and specificity of screening tests and screening uptake rates (Gillespie et al., 2011).

Results

The costs incurred per GDM diagnosis varies by country, and is based on the unit cost of the test technique, the screening strategy and GDM prevalence in the population. The cost per case of GDM detected is calculated as AUS $10 using ADIPS criteria (Moses et al., 1997). With high and low risk factor screening based on criteria established by the ADA, in Iran the cost of screening for GDM was reported to decrease from US$3.80 to US$3.21, and the cost per detected case of GDM declined from US$80.56 to US$77.44 (Larijani et al., 2003). In Italy in 2002, a cost per case of GDM detected of €6.23 was generated using 50g GCT with 100g OGTT. In the same study, the costs per case management were €366.40 (Di Cianni et al., 2002). In the USA the total cost per case diagnosed by universal screening was reported to be US$2,292 (Lemen et al., 1998). Gillespie and colleagues estimated the total health care cost in Ireland to be €12,433,320 and the average cost per case detected was €1,621 (Gillespie et al., 2011). Moreover, in the same study a new estimate of the cost of universal screening for GDM recommended by the International Association of Diabetes and Pregnancy Study Group in Ireland was €46,311,301. The average cost per case detected and the average total cost per case detected and treated were €351 and €9,325, respectively.

Many reports of cost analysis for GDM have determined the costs of screening and diagnostic strategies, but have not incorporated data on the costs of management during pregnancy or clinical adverse outcomes. A cost-consequence analysis on hospital treatment with insulin among 100 women with GDM revealed that the incremental cost of perinatal deaths prevented was $60,506 and life year gained was $2,988 (Moss et al., 2007). However, actual management care costs will depend on the specific composition of the strategies employed in each study.

Table 6.4 ***Summary of cost analysis studies***

Author	Country / (Cost /years)	Interventions	type	Cost data sources	Outcomes	Results	Result (SA)
(Moses et al., 1997)	US (AUS dollar /1995)	Universal screening by ADIPS criteria	CA	Cost estimated	Case detections	Per case detection around $10.00	NR
(Lemen et al., 1998)	USA (US dollar/NR)	Universal screening by ADA criteria	CA	Cost estimated (direct cost)	Case detections	Per case diagnosed $2,292	NR
(Di Cianni et al., 2002)	Italy(Euro/2002)	S1: 50 g GCT + 100 g OGTT S2: Selective screening	CEA	the reimbursement tariffs of the Italian public health	Case screened	Per case screened was € 6.26. Per management case of GDM was € 366.04.	NR
(Scott et al., 2002)	UK/(Pound/NR)	Systematic review	SYS	Cost estimated(direct cost)	Screening test	Total cost of screening (50-g GCT)£8.40, (FPG)£4.10, (75g OGTT)£19.80.	NR
(Larijani et al., 2003)	Iran (US dollar/2002)	S1: High risk by ADS Criteria S2: Intermediate risk S3: Low risk	CA	NR	Case detections	Per case detection **S1:** (130mg/dL) $ 80.56, (140mg/dL) $77.44. S2: (130mg/dL) $66.88, (140 mg/dL) $ 65.63	OSA: Sen / Spe of tests
(Moss et al., 2007)	AUS(AUS dollor/NR)	S1: Intervention groups S2: Routine groups	CCA	Hospital costs (direct out and in patient)	Out-patient service	Cost of outpatient service in S1: $674 , S2 $496. IC: perinatal death prevent $60,506 IC: life year saved $2,988	NR
(Gillespie et al., 2011)	Ireland (Euro/2008)	Universal screening	CA	Cost estimated	Health care cost of GDM	Total health care costs of GDM was estimated at €12,433,320	OSA: prevalence, Sen and Spe of tests
(Agarwal et al., 2012)	United Arab Emirates (US dollar/NR)	S1:50g GCT+ 100g OGTT S2:75g OGTT S3. FPG + 100g OGTT	CA	The laboratory workload units (WLU)	One year cost of GDM	One year cost for (S1) $31,985, (S2) $55,250, (S3) $35,875. WLU for (S1) $28975, (S2) $18,662, (S3) $12,215.	NR
(Gillespie et al., 2012)	Ireland (Euro/2010)	Universal screening	CA	Cost estimated	Case detections	Total health care costs of GDM was estimated at €46,311,301 Total cost per case detected €9,325	NR

NR= Not report, ICER= Incremental cost- effectiveness ration, SA= Sensitivity analysis, OSA= One way sensitivity analysis, CA= Cost analysis, CBA= Cost benefit analysis, CMA= Cost minimisation analysis, SYS= Systematic review. QALY= Quality adjusted life years, DALY = Disability adjusted life years, IC= Incremental cost.

6.6 Review of the cost-effectiveness study

Over the past decade, eight economic evaluation models of screening tests for GDM have been published. Cost-effective analysis and cost-utility analysis studies are methods which form part of a complete economic-evaluation, with the aim of comparing the costs and outcomes of various measures, and are discussed in this section. The summary of all studies is shown in Table 6.6.

Screening test strategy

The studies in this review employ a range of different screening test strategies; two use IADPSG (Werner et al., 2012) (Mission et al., 2012), another two studies use both WHO and O' Sullivan strategies (Poncet et al., 2002) (Marseille et al., 2013), one study uses ADA (NICE, 2008b). The rest of the studies use screening test strategies other than those listed above (Nicholson et al., 2005) (Round et al., 2011) (Lohse et al., 2011). These guidelines recommend the same period of time to screen pregnant women, between 24 and 28 weeks of gestation, for the diagnosis of GDM but recommend different test techniques and test approaches.

Methods

All studies used the decision tree to graphically model the consequences of screening test results and outcomes, and adequately reflect the time dependence of events (Poncet et al., 2002) (Nicholson et al., 2005) (NICE, 2008b) (Round et al., 2011) (Werner et al., 2012) (Lohse et al., 2011) (Mission et al., 2012) (Marseille et al., 2013). The basic design of all 8 models in the review displayed the possible consequences of screening for and treating GDM in decision trees based on different screening strategies, as shown in Table 6.5. The structure of decision trees, for short term complications, common to all the studies in this review, start with a decision node for test strategies followed by chance nodes for true disease state, post-test classification, antenatal and perinatal intervention, perinatal and offspring adverse outcomes (Poncet et al., 2002) (Nicholson et al., 2005) (NICE, 2008b) (Mission et al., 2012). The structure of the decision trees for long term complications were based on the structure used for short term complications and in two studies was followed by intervention to prevent type 2 DM (Lohse et al., 2011) (Round et al., 2011), while in one other study this structure was followed by a postpartum screening test and intervention to prevent type 2 DM (Werner et al., 2012).

Round et al was the only study that presented the perspective in terms of third party payer (Round et al., 2011). Two articles were investigated from the societal perspective (Poncet et al., 2002) (Nicholson et al., 2005).

Four studies of the economic evaluation of GDM employed “no screening” as the reference case. This is the most widely used comparator for diagnostic test evaluation (DRUMMOND, 1987) (Nuijten et al., 1998) (Husereau et al., 2013). Of the eight studies, one study that developed a model to asses CEA for GDM did not mention a comparator in the model (Lohse et al., 2011).

Cost data mostly comprised of direct costs, including the cost of screening and diagnostic tests, antenatal care for mothers with GDM (diet and exercise counseling, glucose control medications, and treatment for complications, were derived from published studies or national cost report (Nicholson et al., 2005) (NICE, 2008b) (Round et al., 2011) (Werner et al., 2012) (Mission et al., 2012). In a study in France, the cost components were calculated by using a prospective study of 120 pregnancies, which included costs of the screening tests, costs of the obstetrical follow-up, costs of the management of GDM, and costs of delivery care (Poncet et al., 2002). Two studies used data on costs that were collected from available data in each setting, and the literature (Lohse et al., 2011) (Marseille et al., 2013). The entire clinical data presented in the papers were obtained from literature reviews based on similar population or databases of each country. Additional input data were also obtained through expert opinion. Clinical model parameters included: the probability of adverse outcomes in both mothers and offspring, probability of GDM treatment, health related quality of life mother and offspring, and test sensitivity and specificity. Various choices of clinical health outcomes in both mothers and offspring have been used in the economic evaluation of GDM, which are detailed in Table 6.5. The outcomes of the relevant parameters relate to maternal adverse complications, neonatal adverse complications, post-partum intervention and long term complications of type 2DM.

Measures of changes in health related quality of life that contributes to QALY or other preference-based measures have been used in the studies in this review. Only five of the eight CEA used QALY as outcome measures for the economic evaluations of screening tests for mothers with GDM (Nicholson et al., 2005) (NICE, 2008b) (Round et al., 2011) (Werner et al., 2012) (Mission et al., 2012). One study used per-case prevention rates as natural units (Poncet et al., 2002), and the last two used averted disability adjusted life-years (DALY) (Lohse et al., 2011) (Marseille et al., 2013). One of the studies did not state the discount rate for both costs and utilities in the study (Poncet et al., 2002).

Two papers in the USA performed a 3% discount rate (Nicholson et al., 2005) (Werner et al., 2012). A 3.5% discount rate was used in the UK (Lohse et al., 2011) .

Five studies considered uncertainty within their analyses, with one-way sensitivity analysis being used in five of the eight papers (Poncet et al., 2002) (Nicholson et al., 2005) (Werner et al., 2012) (Mission et al., 2012) (Marseille et al., 2013) . Additionally three of the eight papers used probabilistic sensitivity analysis (Round et al., 2011) (Lohse et al., 2011) (Werner et al., 2012).

Table 6.5 Details of the decision tree models for GDM

Cost and clinical parameters	(Poncet et al., 2002)	(Nicholson et al., 2005)	(NICE, 2008b)	(Round et al., 2011)	(Lohse et al., 2011)	(Werner et al., 2012)	(Mission et al., 2012)	(Marseille et al., 2013)
Screening test	- Risk factors - Screening test - Diagnostic test	- Risk factors - Screening test - Diagnostic test	- Risk factors - Screening test - Diagnostic test	- Risk factors - Screening test - Diagnostic test	- Screening test - Diagnostic test	- Screening test - Diagnostic test	- Screening test - Diagnostic test	- Screening test - Diagnostic test
Antenatal Treatment intervention	- Endocrinology consultations - Diet - Biological test - Drugs	NR	-Diet -Pharmacotherapy (Glibenclamide, metformin) - Insulin therapy	- Blood monitoring - Insulin therapy	Treatment and other interventions	- Nutritional counseling - Blood glucose monitoring - Insulin therapy	NR	Treatment and other interventions
Maternal Adverse complications	- Prinatal morality - Hypertensive disorders	- Hypertensive disorders - Polyhydramnios - Cesarean section and vaginal delivery with complications	- Caesarean section -Induction of labour	- Caesarean section	Perinatal adverse outcome death mother	- Preeclampsia - Cesarean section - Preterm birth	- Preeclampsia - Cesarean section - Maternal death	Perinatal adverse outcome death mother
Neonatal Adverse complications	- Macrosomia - Prematurity	- Macrosomia - Mild hypoglycemia - Respiratory distress syndrome - Shoulder dystocia - mild, moderate, severe in morbidity	- Neonatal death - Bone fracture - NICU - Shoulder dystocia -Jaundice	- Hypoglycemia - Neonatal death - Shoulder dystocia - Nerve palsy - Phototherapy - NICU	Perinatal adverse outcome death baby	- Shoulder dystocia - NICU - Still birth	- Macrosomia - Shoulder dystocia - NICU - Still birth - Respiratory distress syndrome - Shoulder dystocia -Hyperbilirubinemia	Perinatal adverse outcome death baby
Post partum intervention	NR	NR	NR	NR	- Life style intervention	- Screening - Intensive exercise - Nutritional counseling	NR	- Life style intervention
Long term complication(T2DM)	NR	NR	NR	NR	Mother and child	Mother	NR	Mother and child

Results

Evaluating the cost-effectiveness of universal screening and selective screening versus no screening of the population, a French study reported that the screening of all pregnant women by the WHO strategy with 75g OGTT was the most efficient strategy and selective screening of pregnant women with high risk factors with 50g OGTT had the most favorable ICER (Poncet et al., 2002). Similar conclusions were drawn from studies that evaluated various screening tests against “doing nothing”. In the UK, evaluating the universal screening of all pregnant women and selective screening compared with no screening, from the analysis it would seem that offering a diagnostic test to women from a high-risk ethnic background would be cost-effective compared to not offering screening, with an ICER of £3,677(NICE, 2008b).

Nicholson and colleagues showed that the combination of 50g GCT and 100g OGTT (S1) was the most cost-effective strategy when ICER of < $50,000/QALY were considered highly favorable toward a particular screening strategy (Nicholson et al., 2005). Similar conclusions were drawn from other cost-effectiveness analyses which stated that the combination of a screening test with a diagnostic test or a two-step approach was cost-effective. For example, in the USA a study showed that the IADPGS approach for GDM screening tests (fasting plasma glucose) combined with a diagnostic test (75g OGTT) was cost-effective. That is to say, compared to the no-screening strategy or the current screening strategy, providing the GDM diagnosis gives an opportunity for early and intensive intervention and the prevention of future overt DM (Werner et al., 2012). However, if a one-step approach to screen pregnant women, with ceiling ratios ranging from 50,000 – 100,000/QALY was used, screening with the OGTT (diagnostic test) was more expensive and cost effective at $61,503/QALY, than the 50g GCT screening test (Mission et al., 2012).

However, only one study, in England, that evaluated the cost-effectiveness at different prevalence’s in the population reported that the most preferred option was the most effective strategy within the maximum willingness-to-pay threshold (£20,000 per QALY in this case). However, if the GDM risk was < 1% then the no screening/ treatment strategy is cost-effective. Where risk was between 1.0% - 4.2%, combinations of fasting plasma glucose (FPG) and 75g OGTT are the most cost-effective, and where risk was >4.2 screening all women with OGTT was cost-effective. This study concluded that screening strategies were most likely to be cost-effective, based on the range of GDM risk factors (Round et al., 2011).

Results of the study undertaken by Lohse and colleagues stated that net savings per woman for universal screening followed by postpartum lifestyle management in India was $78 and in Israel it was $1945. Cost-effectiveness DALYs averted in India were $256 and in Israel $2584. Nevertheless, the current findings of cost-savings or favourable cost-effectiveness are robust to a wide range of plausible input values (Lohse et al., 2011). Similar economic evaluations of screening tests for GDM were conducted in the same setting within the real data. In terms of estimated costs and effectiveness of screening and treating mothers with GDM, considering PAEs and type 2 DM, ICER per DALY averted in India and Israel were $1,626 and $1,830, respectively. In both settings, cost-effectiveness was sensitive to the incidence of type 2 DM and to the costs and effectiveness of postpartum intervention (Marseille et al., 2013).

Table 6.6 Summary of eight cost-effectiveness studies

Author	Perspective	Country / (Cost /years)	Interventions	Discount rate	Outcome	Results (ICER)	Results (SA)	Grading
Poncet et al (Poncet et al., 2002)	Societal	France (Euro/ 2001)	S1: high risk + 50g OGTT S2: 50g OGTT S3: 75g OGTT S4: No screening	NR	ICER: per-case prevention rates (macrosomia, prematurity, perinatal mortality, and hypertensive disorder)	- S2and S1: additional effectiveness 1.10-1.11 - S3and S1 time additional effectiveness 3.27- 3.75 time	OSA: Results most sensitive to (i) Cost	++
Nicholson et al (Nicholson et al., 2005)	Societal	USA/ (US dollar / 2003)	S1:50g GCT + 100g OGTT S2:100g OGTT S3:75g OGTT S4: No screening	3% in both cost and benefit	ICER : life of year saved QALY for maternal	Reference $ 32,374 Dominated Dominated	OSA: NR	+
			S1:50g GCT + 100g OGTT S2:100g OGTT S3:75g OGTT S4: No screening	3% in both cost and benefit	ICER : life of year saved QALY for neonatal	Reference $ 8,252 Dominated Dominated		
NICE (NICE, 2008b)	NR	England and Wales / (Pound sterling / 2008)	S1: OGTT S2-S5: FPG,RBG,GCT,FPG +75g OGTT S6: OGTT ADA criteria S7:GCT ADA criteria S8-S10: FPG,RBG,GCT S11-S12: FPG,GCT + 75g OGTT S14-S18: age≥ 25 and30 FPG,GCT + 75g OGTT S19-S21:high-risk ethnicity FPG,GCT +75g OGTT	NR	ICER : QALY gained	-S21(high risk ethnicity + OGTT) and no screening: ICER £3,677 - S6 and S21: ICER £ 21,738(higher than the £20,000 per QALY threshold suggested by NICE)	PSA: prevalence, treatment options, and the efficacy of using risk factors to define high and low risk population	++

NR= Not report, ICER= Incremental cost- effectiveness ration, SA= Sensitivity analysis, OSA= One way sensitivity analysis PSA= Probabilistic sensitivity analysis, QALY= Quality adjusted life years, DALY = Disability adjusted life years.

Table 6.4 (Continued)

Author	Perspective	Country / (Cost /years)	Interventions	Discount rate	Outcome	Results (ICER)	Results (SA)	Grading
Round et al (Round et al., 2011)	third-party payer (NHS)	England and Wales / (Pound sterling / 2009)	S1: 75 g OGTT S2:FPG S3:RBG S4: 50g GCT S5: PBG + 75g OGTT S6: FPG + 75g OGTT S7: 50g GCT + 75g OGTT S8: No screening	3.5% in only benefit	ICER : life of year gained QALY over life time	No QALY present in the paper.	PSA: Present which treatment more likely to be cost-effective by a range of GDM risk	++
Lohse et al (Lohse et al., 2011)	NR	India and Israel / (US dollar / 2011)	S1:Universal screening	5% in costs, 3% for all future costs	ICER : adverted disability-adjusted life-years (DALY)	DALYs adverted: India: $256, Israel: $ 2584	PSA: robust to a wide range of plausible in put value, including highly unfavourable values.	++
Werner et al (Werner et al., 2012)	Health care provider	USA / (US dollar / 2011)	S1: no screening S2: 50gGCT +100g OGTT S3: FPG + 75g OGTT	3% in both cost and benefit	ICER : life of year gained QALY over life time	Reference $ 16,689 $ 20,336	OSA: long-term behavioural intervention in GDM PSA: S3 CE with Probability 96.4 %	++
Mission et al (Mission et al., 2012)	Societal	USA / (US dollar / 2012)	S1:50gGCT S2:75g OGTT	3% in only benefit	ICER : QALY gained	Dominated $61,503	OSA: Percentage of additional GDM diagnosed with 2h OGTT, cost of 2 h OGTT, Cost to treat GDM, and Efficacy of treatment	++
Marseille et al (Marseille et al., 2013)	NR	India and Israel / (US dollar / 2011)	India S1: no screening S2: 75g OGTT	3% in costs	ICER : adverted DALYs and net cost per DALY	Reference $1,626	OSA: the estimate incident of Type 2 DM in mother, cost per partum care.	++
			India S1: no screening S2: 75g OGTT	3% in costs	ICER : adverted DALYs and net cost per DALY	Reference $1,830		

NR= Not report, ICER= Incremental cost- effectiveness ration, SA= Sensitivity analysis, OSA= One way sensitivity analysis PSA= Probabilistic sensitivity analysis, QALY= Quality adjusted life years, DALY = Disability adjusted life years.

6.7 Discussion

This is a systematic review of the literature focusing on the costs and cost-effectiveness analysis of GDM. Previous to this review, available data in this field of research had been summarised unsystematically. Cost analysis was separated from the economic evaluations (CEA, CUA) because cost analysis is not a full economic evaluation (Drummond, 1987). However, the cost analyses do give important information on the key costs associated with screening tests for GDM. In this review, 9 studies in cost analysis and 8 studies in cost-effectiveness analysis were identified.

6.7.1 Quality assessment

All economic evaluation studies that were included in this systematic review were good in terms of their overall quality; Seven were graded as "++", indicating very good quality with only one study being graded as "+", indicating moderate quality. Overall, the studies performed well in terms of study design. The studies met the criteria of stating the research question and describing the alternative screening test that they used to compare their evaluations. These eight studies also clearly detailed costs and price adjustments. Most of the studies met the criteria of stating their source material for the estimation of treatment effectiveness, reporting the results to the answers of each of the research questions. Almost all of the studies accounted for the costs and benefits and then discounted them from their results. Two short term studies did not apply discount rates for their costs and benefits.

Some limitations could be found for the following criteria; more than half of the studies reviewed did not report the difference in cost and outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics. However, this is perceived to be a minor limitation only, which would not affect the overall quality of this study.

6.7.2 Cost analysis studies

Costs associated with the screening and management of GDM vary widely by country and are dependent on tests used (FPG, 50g GCT, 75g OGTT and 100g OGTT), screening approach (universal and selective), screening steps (one-step and two-step), diagnostic criteria and prevalence of GDM in the population.

In the early part of the study period for the review of cost analyses of GDM, studies reported cost per case detection at similar levels. Many of the reports have determined outcomes expressed in natural units, but few studies incorporated data on the cost of management during pregnancy or clinical adverse outcomes. In this early part of the study period, authors only pay attention to the cost of screening tests. However, results of cost analysis do not give much information in order to compare the screening test strategy, therefore, nowadays, decision-analytic models that represent a technique to synthesise evidence that is currently available on the costs and consequences of alternative health care interventions have become more important. Both screening test and treatment data from developing countries are not present in the published literature.

Given those findings, two studies that compared screening strategies stated that universal screening is less expensive than selective screening (Di Cianni et al., 2002) (Larijani et al., 2003). These studies suggest that selective screening may be an option in a situation where healthcare is very limited. Moreover, a selective approach tends to be performed in populations with a low prevalence of GDM; hence, there would be a small number of cases missed by selective screening. There is only one study that presents incremental costs (Moss et al., 2007). The treatment of mothers with GDM showed the reductions in incremental costs per perinatal death prevented and serious perinatal complication prevented, which was compared with routine service care. Moreover, the incremental cost per extra life year gained is highly favorable.

The costs of screening test techniques in this review show that screening tests, (FPG and 50g GCT), are less expensive than diagnostic tests (75g OGTT and 100g OGTT) (Scott et al., 2002). No previous studies have been conducted for costs in the UK. However, Scott and his colleagues conducted a systematic review and an economic evaluation of GDM in 2002 (Scott et al., 2002). In The Scott study, ten published articles on the costs of screening tests for GDM were reviewed between 1983 and 1998. Two of Scott's studies were included in this systematic review. For example, Scott and colleagues estimated the costs of screening tests in the UK, including the costs of material used in the various screening and diagnosis tests, staff time for performing the tests, and biochemical analysis. They concluded that the total cost of FPG, 50g GCT and 75g OGTT were £4.10, £8.40 and £19.80, respectively (Scott et al., 2002). However, in the absence of any published studies or locally available data for UK costings, Scott's study calculated screening costs using data generated by the Scottish Health Purchasing Information Centre's (SHPIC) costing unit. No economic evaluation modeling was reported in Scott's study.

6.7.3 Cost-effectiveness analysis studies

Decision-analytic models represent a technique to synthesise evidence that is currently available on the cost and consequence of alternative health care interventions (Philips et al., 2004). The use of decision-analytic modeling for health technology assessments has seen significant use in recent years. New diagnostic technologies have the potential for incurring substantial additional expenditure, creating the need for comparative economic and clinical analysis. Therefore, economic assessments of diagnostic tests are necessary for health technology assessment (HTA).

This systematic review identified 8 CEA. Over the past decade, decision making in healthcare has increasingly been based on cost-effectiveness considerations. This trend is reflected in most national guidelines concerning GDM screening. Many national guidelines have recommended screening test strategies for GDM, such as O'Sullivan and Mahan (O'Sullivan and Mahan, 1964), The National Diabetes Data Group (NDDG) (National Diabetes Data Group, 1979), The American Diabetic Association (ADA) (American Diabetes Association, 1998), The World Health Organization (WHO) (WHO, 1999a), The National Institute for Health and Clinical Excellence (NICE) (NICE, 2008d), and The International Association of Diabetes and Pregnancy Study Group (IADPSG) (Metzger et al., 2010b). All studies in this review screen pregnant women during the same gestational time period, between 24 and 28 weeks. The studies in this review used different screening test techniques, such as RPG, FPG, and 50g GCT, and different screening test approaches, either universal and selective screening or one–step or two-step approaches. However, all studies used 75g OGTT and 100g OGTT as standard tests and the assumed sensitivity and specificity of those tests is 100%. It is crucial to select the appropriate comparators for the analysis of CEA (NICE, 2013). "No screening" or "doing nothing" is suitable as a comparator for screening tests for GDM as it informs decision makers of the full impact of the alternative screening tests.

This systematic review has shown that the most common decision model used to evaluate cost-effectiveness analyses of different screening test strategies for GDM is the decision tree. The simple decision tree is usually appropriate for situations where the time frame is short or when the intervention does not change the underlying disease process, as with screening tests that can only detect the presence of a disease (Barton et al., 2004). While a Markov model is not inappropriate, it may be unnecessary where this is true. Decision trees for screening tests are unlike those for chronic disease which can become very bushy in modeling, making them unwieldy and complex to programme and analyse (Drummond, 1987) (Drummond and McGuire, 2007). It is therefore appropriate to use the decision tree to structure a decision model for screening tests for

GDM in both short and long term complications. Decision tree models are also suitable for the decision problem of screening for GDM as they can give a detailed description of the disease process (Philips et al., 2004). Decision trees for screening tests estimate the proportion of people tested and diagnosed and the testing costs for each screening strategy. These estimates applied to a population cohort, allow the model to estimate expected costs and expected outcomes. Any limitations of the decision tree models can be resolved using Markov models, which deal with patterns of recurring–remitting GDM after postpartum (GDM in subsequent pregnancy) and long term adverse complications with type 2 DM over a period of time.

Information on populations for which the cost-effectiveness of screening tests for GDM that are to be assessed is determined by the respective countries where the studies are performed. It is assumed that countries have their own national estimates on prevalence for specific disease areas and for population sub-groups. Costs and clinical parameters for short and long term complications were taken from published studies. All studies in this review stated the decision-making perspective of the study clearly and included all relevant costs for each study. The results of economic evaluations are likely to be dependent on the perspective employed and should be stated and defined (Roberts et al., 2012a).

The earlier studies of economic evaluations for GDM assessed cost and effectiveness in short term complications in mothers or offspring whereas three of the four later studies built models that included postpartum intervention to prevent DM in mothers (Werner et al., 2012), as well as in both mothers and offspring (Lohse et al., 2011) (Marseille et al., 2013). Variability was observed in terms of what costs and outcomes were modeled in each of the studies. The economic evaluations of screening tests for GDM should consider both short and long term complications. Case identifications and short term complications are not the endpoint outcomes for GDM screening tests. The effect of screening and the treatment of GDM, that includes blood glucose monitoring, oral medication and regulating insulin, is a reduction in the likelihood of developing short term complications and type 2 DM in both mothers and offspring. Long term consequences are appropriate outcomes for the CEA of screening tests for GDM. A major reason for introducing screening programmes in the population is that it is important to initiate the appropriate treatment after positive cases have been identified. When patients have positive screening test result, this may reduce their quality of life. This can only be compensated for by providing treatment that improves health and prolongs life. Therefore, using screening tests for GDM over multiple time horizons, such as one year and over a lifetime, might be appropriate for the analysis of data with respect to short and long term complications.

One study used per-case prevention rates as natural units and did not apply a discount rate (Poncet et al., 2002). This was because the model measures outcomes in terms of short term adverse complications, and due to the short time frame, costs and outcomes do not need to be discounted. In the UK, NICE suggests a discount rate of 3.5% for both costs and outcomes (NICE, 2013). Models, for which the costs and outcomes of screening are not realised for several years however, need to be discounted. Discount rates tend to vary depending on location, setting, time horizon and perspective of the analysis (Husereau et al., 2013).

Two outcome measures were used, including QALYs and DALY, in the economic evaluation of screening tests for GDM. Health related quality of life measures (QALYs) are recommend for use as a measurement of health outcomes to ensure that cost-effectiveness results are comparable to some extent between countries. Round and colleagues used QALY during pregnancy and three months after deliver for short term adverse complications for randomised clinical trials, which used short form – 36 (SF-36) (Crowther et al., 2005). SF- 36 is not an appropriate measure of health outcomes in CEA because it is not a preference-based measure of health-related quality of life (HRQoL). The generic preference based instrument with EuroQul- 5D (EQ-5D) is a popular method usually used for measuring changes in health-related quality of life in CEA studies (Richardson and Manca, 2004) (Husereau et al., 2013). The preferred outcome measure in economic evaluations is QALY, which combines length of life and HRQoL in a single metric (Husereau et al., 2013), and is considered a particularly useful outcome measure for economic evaluations of screening tests for GDM. QALY can also be compared both within and across disease areas. As methodologies used in cost-effectiveness studies have evolved differently, it would be beneficial to use similar outcome measures such as QALY to increase the comparability of studies. The use of a range of different outcome measures can cause problems for decision makers, who have to make judgements based on disparate results.

Two studies in this review measured health outcomes in terms of DALY, for both short and long term complications. The DALY for short term complications were estimated based on health status utility. Whereas, for long term complications they were estimated based on lifetime utility of type 2 DM in mothers with a history of GDM, while for offspring they were obtained by comparing disability adjusted life expectancy with and without type 2 DM at the estimated age of onset of type 2 DM (Lohse et al., 2011) (Marseille et al., 2013). Both QALY and DALY are based on preferences of health and are appropriate to measure health outcomes in CEA (Marseille et al., 2013).

In offspring, various health outcomes and techniques were used to estimate health utility. Two studies retrieved data from the literature and measured outcomes in terms of categories of health states such as normal health (utility 1), preterm birth (utility 0.96), permanent brachial plexus injury (utility 0.87) and death (utility 0) (Nicholson et al., 2005) (Werner et al., 2012). Other papers in this review estimated (by weighting average) QALY for serious perinatal complications, calculated based on the relative frequency of each individual component. Outcomes used in economic evaluations for offspring are more difficult to estimate than for adults. The choice of outcomes affects the findings of an economic evaluation, and so it is important that the reasons for choosing one measure of outcome over any other should be outlined (Husereau et al., 2013). Because the studies used different costs, outcome measures and made different assumptions about the parameters used, it was not possible to draw comparisons between the results of the studies. Also different strategies in modeling and different outcomes used were not allowed for comparison.

The differences in effectiveness, and hence cost-effectiveness, among the models is related to the different assumptions each model makes about the effectiveness of screening tests for GDM. Therefore this review cannot definitively state the single most-effective or cost-effective strategy for screening tests for GDM. Four of the studies compared various screening test strategies with risk factor screening (selective screening) in the study (Poncet et al., 2002) (Nicholson et al., 2005) (NICE, 2008b) (Round et al., 2011). Two of those studies stated that the most favourable ICER was the screening of high-risk pregnancy with a screening test technique such as 50g OGTT (Poncet et al., 2002) and 100g OGTT (Nicholson et al., 2005). Two studies in the UK suggested the optimal strategy varies according to the women's risk factors (Round et al., 2011) (NICE, 2008b). This review can say that in the low risk factor group, the most effective strategy is to do nothing. Screening that is based on individual risk factors is possibly more sophisticated and cost-effective in comparison to methods based on multiple risks that are commonly considered in the literature. As only a few of the studies in this review combine screening test techniques and risk factor screening it was not possible to specify what the most appropriate screening test strategy should be.

Furthermore, none of the models address the potential QALY gains of screening in terms of subsequent pregnancies or delayed progression or developing type 2 diabetes. One potentially essential outcome for the detection of GDM is the identification of women who are at high risk of developing type 2 DM or at high risk of GDM in subsequent pregnancies. The model for economic evaluation for GDM should account for the effectiveness of postpartum screening for type 2 DM. In addition, there were no studies that presented the results of value of information (VOI) for screening tests for

GDM. VOI is used to make meaningful recommendations to decision makers, who can then make informed decisions as to whether a new intervention should be adopted or reject based on current evidence, or whether further information is required to help make the decision.

A willingness to pay threshold exists in some countries, but not for all studies. An interpretation of whether a certain screening test is cost-effective or not can only be made if there is a monetary value that is deemed appropriate to achieve one additional QALY. However, if willingness to pay is not available, the conclusion after modeling different strategies can only be in relation to how much it would cost to gain one QALY and not whether it would be cost-effective to implement a particular strategy compared to another.

The interpretation of any cost-effectiveness result will depend on the level of uncertainty that surrounds the input parameters. A series of one-way and multi-way sensitivity analyses depicted what would happen when the value of one or more of the parameters were changed, including prevalence, treatment options, and the efficacy of using risk factors to define high and low risk populations. An assessment of uncertainty should be included within an economic evaluation to reflect costs and health outcome results, which were obtained from the study. Prevalence, costs and the efficiency of using risk factors to define high and low risk population parameters are the greatest influence on a model's results.

6.7.4 Limitation and recommendation for future research

This systematic review ignores the critical appraisal of cost analysis papers. The study of cost analysis, which is a partial form of economic appraisal, only looks at the costs of the programs and does not provide information on the health outcomes of interest. Presently there is no accepted criterion for the quality assessment of cost analysis studies available. It is difficult to assess the quality of cost analysis studies.

6.8 Conclusion

This chapter has presented a review of published papers about the economic evaluation for GDM and has critically appraised the economic evaluation of screening tests for GDM. A systematic review of all publications has been conducted that has addressed the health economic evaluations of screening tests for GDM, as well as their structural and methodological aspects.

The review of the CEA study has significant implications for future research and policy making. The available information on cost effectiveness provide strong evidence in support of the use of risk factor screening tests for GDM in high-risk groups, such as family origin with a high prevalence of diabetes, previous gestational diabetes and obesity. The difference within the reviewed studies in terms of: study design, comparators, interventions, outcome measures and the analysis of uncertainty makes meaningful comparison between the studies very difficult. A decision must be made as to the most suitable screening test based upon and supported by all available knowledge.

Based on these findings, the following two chapters detail and construct an economic evaluation for GDM, which may estimate the benefit of diagnostic tests as a function of their expense and the prevalence of disease. The decision tree for case identification will be developed in the next chapter based on its findings.

Chapter 7 Decision analysis for screening tests for GDM: Case identification

7.1 Introduction

The aim of this chapter is to use decision analytic modeling in order to compare the relative costs and outcomes (effects) of five strategies of screening programmes for GDM. Health economists use economic models to represent real systems where information on the natural history of disease is combined with measures that quantify the major drivers of cost and effect. Decision makers can make reimbursement decisions based on all the available information relating to the disease and the effectiveness of its treatment. In addition, real systems are highly complex and economic modelling provides a way in which one can combine information on the natural history of disease and the effect of treatment in a potentially useful and meaningful way. Once these elements are clearly understood, then an appropriate modelling structure can be selected which allows for these relevant elements to be included within the model. With structural decisions made, the model can be developed using available and relevant data sources.

This chapter begins by identifying the decision problems of screening tests for GDM in Scotland. The subsequent section then illustrates the structure of a decision tree that involves five different strategies for screening tests in Scotland. There then follows a discussion on how to apply the decision tree analysis to the evaluation of diagnostic tests for GDM. Within the decision tree, different test methods are applied, dependant on the strategies employed. In the section following this all costs and outcome parameters are presented and discussed and then the final section of this chapter cost per case identifications are calculated and presented in terms of both deterministic and probabilistic analysis dependant on the screening strategy.

7.2 A new economic model for GDM screening tests

The majority of previous full economic evaluation models have used the decision tree to structure economic modelling for GDM screening tests. The decision tree is probably the most common structure that describes a clinical decision and its possible outcomes (Drummond, 2005a) (Briggs et al., 2008). The basic principle of the decision tree in economic evaluations is detailed in Chapter 2 section 2.7.5. The decision analysis

method has been proposed to evaluate screening and diagnostic tests in terms of possible clinical consequences; occurrence or no-occurrence of an event. The statistics of the diagnostic outcome basically consist of the outcome of the test (T) and the presence of disease (D). The decision analysis measures the association between the tests (T), with positive (T+) and negative (T-) results and the presence of disease (D): with disease (D+) and without disease (D-) (Andrew, 2008).

In previous decision modelling of GDM simplified decision tree models were based on the screening strategy, disease status and screening test accuracy, and dependant on the strategies and studies discussed in Chapter 6. In this model, decision trees screening tests for GDM start with screening strategies, similar to the decision trees in other economic evaluations of GDM. Following the screening strategies, chance nodes that present true disease status (prevalence) are then followed by screening and diagnostic test classifications such as true positive (TP), false negative (FN), false positive (FP), and true negative (TN). At this point, the decision trees can indentify mothers with and without GDM.

The decision trees in other economic evaluations of GDM, as reviewed in chapter 6, were built based on the screening tests used in each strategy, antenatal treatments, short term complications in mothers or offspring, and long term complications in mothers and offspring. In this model, following the screening test strategy (case identification) section, the decision tree is broken down into three separate decision trees; a decision tree for antenatal treatment, decision trees for both short and long term adverse complications. These three individual decision trees are attached to the screening test decision tree as mentioned above and are outlined in the following chapter.

In conclusion, within this chapter a new economic model for GDM screening test case identification in Scotland is developed and detailed. The model is constructed similar to decision trees in previous studies while avoiding the limitations of economic evaluations in previous studies as mentioned in the preceding chapter.

7.3 Decision making in screening tests for GDM

Economic evaluations have been applied in different areas of medical and public health programmes, such as primary prevention (risk factor reduction), secondary prevention, screening and diagnostic test programmes. In diagnostic procedures, clinical decision analysis requires a model to link the diagnostic accuracy data and the long-term effectiveness of available treatment. The expected value of diagnostic tests depends

strongly on prevalence of disease, risk association with the test, diagnostic accuracy, and the relationship between the net benefit of treating truly diseased patients and the net harm of unnecessarily treating non-diseased patients. In an economic analysis, savings from advanced disease prevention must be included in the model along with the diagnostic tests and treatment costs. Specifically, the challenge in diagnostic studies is the definition of the optimal cut-off point, which differentiates between positive and negative test results. There is no study, however, on the economic evaluation of screening and diagnostic tests for GDM in Scotland. This study populated a probabilistic model of decision analysis in order to estimate the incremental cost-effectiveness ratios (ICER) of four alternative strategies versus the strategy “no screening”.

7.3.1 Perspective

In economic modelling it is important to decide at the outset what the end use of the model will be. NICE also recommends that all economic evaluations should be based on UK population preference and this was adopted within this chapter (NICE, 2008c). This study preformed a decision analysis from a health care provider’s perspective to inform policy makers on this important issue by establishing the optimal cost-effective screening and diagnostic tests of GDM. The cost of implementing each screening test and related treatment is derived from the study perspective which includes hospital costs and professional fees for the initial screening tests for risk factors. Health sector perspectives concern the health-related cost and impact on the government and on the private sector. Moreover, the next component that is an important perspective is valuation. An economic evaluation of GDM defines the currency reference that will be used to present the resource expenditure associated with a given cost in “pound sterling”.

7.3.2 Interventions and comparators

The model was developed to compare the expected costs and health outcomes of five possible screening strategies of GDM in all pregnant women without a prior diagnosis of diabetes in Scotland, during 24-28 weeks. This study described GDM as carbohydrate intolerance of varying degrees of severity, with onset or first recognition during pregnancy.

This economic evaluation of screening tests for GDM selected screening strategies based on the screening strategy recommendations in the UK list in Table 4.8, and a combination of the test results from section 3.4. Four screening guidelines were selected included SIGN 2001, NICE 2008, Consensus 2010 and SIGN 2010 for the

reasons outlined below and based on an experts suggestion (Robert Lindsay). Table 7.1 shows a summary of the types and methods of screening tests which were used in each strategy. The two guidelines proposed in SIGN 2001 and SIGN 2010 were selected because they were proposed by The Scottish Intercollegiate Guidelines Network (SIGN, 2001) (SIGN, 2010). Although, those two guidelines were published by the same organisation, they recommend different screening test methods and procedures. Although SIGN 2010 is the most recent up-date to this report, it remains unclear as to which of the guidelines is the most effective, as there is no cost-effectiveness study in either the SIGN 2001 or the SIGN 2010. The inclusion of both guidelines in this study therefore allows for a comparison of their relative cost effectiveness. Both of these guidelines have also been compared to the other guidelines referred to in this study.

The third screening test used was the NICE 2008. NICE 2008 published Diabetes in Pregnancy Clinical Guideline 63 which is used throughout England & Wales (NICE, 2008d). NICE 2008 was selected because it is used all over England & Wales and has similar population characteristics to Scotland (Office for National Statistics, 2012) (The Scottish Government, 2011). Lastly, the fourth method selected was the New Consensus Criteria for GDM (Moses, 2010), and was chosen for this study as it is the result of an international consensus, based on the HAPO study as outlined in Diabetes Care (Metzger et al., 2010a). According to the combination of test approaches mentioned in section 3.4, all guidelines that recommend screening tests for GDM employ a negative dominant strategy (NDS).

Moreover, the four screening tests in this model (other than ‘no screening’) include either a one step approach where a diagnostic test is performed without prior screening or a two step approach where a diagnostic test is performed to confirm a positive screening test. In addition, the screening tests use either universal or selective screening procedures. For example, both the NICE 2008 and SIGN 2010 guidelines recommend selective screening using exactly the same risk factors for GDM screening, but different test methods. Universal screening is the screening of all pregnant women during 24 -28 weeks with various tests depending on the strategies recommended in SIGN 2001 and the 2010 Consensus Criteria. Additionally, the various screening test procedures and test methods used in the four different strategies can be matched to the various combinations of test results discussed in section 3.4. Discounting was not applied, as the time horizon was shorter than a year.

Table 7.1 Screening and diagnostic strategies used in cost effectiveness analysis

Strategies	S1:SIGN 2001	S2: NICE 2008	S3:2010 consensus	S4: SIGN 2010
Reference	(SIGN, 2001)	(NICE, 2008d)	(Moses, 2010)	(SIGN, 2010)
Screening type	Universal screening	Selective screening	Universal screening	Selective screening
	2 step approach	2 step approach	1 step approach	2 step approach
Screening threshold	random glucose ≥ 5.5mmol/l	Fasting≥7.0mmol/l	Fasting≥5.1 mmol/l	
Diagnosis threshold	Fasting≥5.5mmol/l OR 2 hour >9.0mmol/l	Fasting≥7.0mmol/l OR 2 hour ≥7.8mmol/l	Fasting≥5.1mmol/l OR 1 hour≥ 10 mmol/l 2 hour≥8.5mmol/l	Fasting≥5.1mmol/l OR 1 hour≥ 10 mmol/l 2 hour ≥8.5mmol/l
Clinical risk factors		- body mass index above 30 kg/m2 - previous macrosomic baby weight 4.5 kg or above - previous gestational diabetes - family history of diabetes (first degree relative with diabetes) - family origin with a high prevalence of diabetes: South Asian, Black Caribbean and Middle Eastern		- body mass index above 30 kg/m2 - previous macrosomic baby weight 4.5 kg or above - previous gestational diabetes - family history of diabetes (first degree relative with diabetes) - family origin with a high prevalence of diabetes: South Asian, Black Caribbean and Middle Eastern

The five different strategies (including 'no screening') used in this model are described in detail in this section in terms of both screening steps and screening methods. Screening test strategy is denoted by "S".

S1: **SIGN 2001**, the screening of all pregnant women with random plasma glucose. The threshold for positive screening is a random glucose value greater than or equal to 5.5mmol/l. Patients with positive random glucose subsequently underwent a fasting plasma glucose test and/or 2 hour 75g OGTT, which was used as the actual diagnosis test for gestational diabetes mellitus. The positive threshold was greater than or equal to 5.5mmol/l or 75g OGTT 2-hour greater than or equal to 9.0mmol/l respectively. This test is recommended by SIGN 2001.

S2. **NICE 2008**, selected pregnant women with one or more specific high risk factors underwent the fasting plasma glucose test, as shown in Table 7.2. A fasting plasma glucose test value greater than or equal to 5.5mmol/l meets the threshold for the diagnosis of gestational diabetes. The patient needs a further diagnosis test. The positive screening test must be confirmed by fasting glucose and 2 hour 75g OGTT for diagnostic test. The threshold for a positive diagnosis is a fasting glucose value greater

than or equal to 7mmol/l and/or a 2 hour 75g OGTT value greater than or equal to 7.8 mmol/l. This test is recommended by NICE 2008.

S3. **2010 Consensus**, all pregnant women underwent fasting plasma glucose or 1 and 2h 75g OGTT screening tests. Gestational diabetes mellitus was diagnosed if the fasting plasma glucose value was greater than or equal to 5.1mmol/l and/or 1 hour greater than or equal to 10mmol/l, 2 hour greater than or equal to 8.5mmol/l respectively. Both screening tests and diagnosis tests used the same threshold. These criteria were proposed by the 2010 consensus.

S4. **SIGN 2010**, all pregnant women with one or more high risk factors were requested to undertake fasting plasma glucose and 1 and 2h 75g OGTT screening tests, as shown in table 3. The threshold for positive diagnosis is fasting plasma glucose greater than or equal to 5.1 mmol/l and/or 1-hour greater than or equal to 10mmol/l, 2-hour greater than or equal to 8.5mmol/l respectively. This strategy was conducted by SIGN 2010.

S5. No screening test, All pregnant women “do nothing” or in other words no pregnant women receive screening or diagnostic tests for GDM, as a comparator intervention. The conditions of “do nothing” represent the standard of care.

Universal screening is the screening of all pregnant women during 24 – 28 weeks with various tests depending on the strategy (S1: SIGN 2001 and S3: 2010 consensus). On the other hand, selective screening is performed during the first interview or visit to a clinic by screening all pregnant women for risk factors. Women who have positive risk factors undergo further screening tests depending on the strategy (S2: NICE 2008 and S4: SIGN 2010). It is assumed that all pregnant women that perform risk factor screening at first visit and undergo screening tests are either positive or negative for risk factors, at 24-28 weeks. The various risk factors are introduced in each of the guidelines, as mentioned in section 4.6.2.1. The most common risk factors referred to in this model are summarised and listed in Table 7.2.

Table 7.2 Clinical risk factors for gestational diabetes mellitus

Clinical Risk Factors for Gestational diabetes mellitus
• Body mass index (BMI) above 30 kg/m^2
• Previous macrosomia baby weighing 4.5 kg or above
• Age over 30 years
• Previous gestational diabetes
• Family history of diabetes (first-degree relative with diabetes)
• Family origin with a high prevalence of diabetes:
– South Asian (specifically women whose country of origin is India, Pakistan or Bangladesh)
– Middle Eastern (specifically women whose country or family origin is Saudi Arabia, United Arab Emirates, Iraq, Jordan, Syria, Oman, Qatar, Kuwait, Lebanon or Egypt).
– Other, South-east Asian, Aborigine, Hispanic, African, Black Caribbean, Pacific Islander and Indigenous Australian ancestry

7.4 Decision tree structure of GDM screening tests: case identification

In all screening test strategies, when the screening test result is positive, the standard procedure is to confirm the result by further diagnostic tests, as mentioned in section 4.7. In other words, only patients whose tests return positive will undergo further testing. In the four screening test strategies used in this study (not including 'no screening test') there also exists the need for further testing when test results are found to be over the accepted threshold, known as the negative dominant strategy (NDS). All guidelines that recommend screening tests for GDM employ the NDS.

In this economic evaluation, decision trees are designed to assess the five strategies of GDM. Decision trees for case identification were built based on the combinations of test approaches that include the negative dominant strategy (NDS) and the positive dominant strategy (PDS) as mentioned in section 3.4. This economic evaluation of GDM screening tests was applied to both of these approaches in order to build decision trees. However, decision tree maps in previous economic evaluations for GDM were applied only to NDS, which offers a second diagnostic test to people who initially test positive in their screening tests.

This model introduces a new approach by applying PDS to the economic evaluation of GDM. This section illustrates and discusses the differences between the decision trees in both NDS and PDS strategies.

As mentioned, the four different screening test strategies in this model (not including 'no screening') have different approaches that include a one step approach, a two step approach, universal screening and selective screening. However in the construction of decision trees for the selective screening approach in previous economic evaluation studies the true disease stage branch precedes the risk factor branch. With such a construction, risk factor screening cannot function as a screening tool. Therefore to improve on previous studies and present risk factors as a screening tool, the decision trees in this economic model begin with risk factor screening for the selective screening approach. Risk factor screening can function as a screening tool and is able to benefit the selective screening approach. The details of the decision tree structure are outlined below.

Decision node: The tree begins with a decision node which indicates the five strategies of the screening test and represents the decision being addressed in the model.

Chance node: In universal screening tests, there are two possible chance nodes: with disease (D+) or without disease (D-). Branches of chance nodes present possible outcomes of probability for the prevalence of gestational diabetes. For selective screening, the initial chance nodes are for risk factor status, considered as risk factor positive (R+) and negative (R-). Following the risk factor chance nodes is the true disease stage with and without risk factors. Subsequently chance nodes indicate the screening and diagnostic stages. Branches of chance nodes illustrate the possible positive (T+) or negative (T-) test results. The series of chance nodes show the following points of uncertainty. The basic statistics for the clinical diagnosis tests are indicated in each branch of possible outcomes of a screening and diagnostic test. Sensitivity is the probability of a positive test result if disease is present, and specificity the probability of negative screening if gestational diabetes is absent.

Terminal node: Each pathway ends in a terminal node, which represents the possible outcome of each pathway. The decision trees terminate at case identifications of GDM with 4 possible outcomes; True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) as shown in Table 7.3. Mothers with TP and FP receive treatment during gestation.

Table 7.3 The details and signs of terminal nodes of screening tests for GDM

Results at terminal nodes	Details
True positive	GDM with treatment
False negative	GDM without treatment
False positive	No GDM with treatment
True negative	No GDM without treatment

7.4.1 Case identification: decision tree structure for GDM screening tests in the negative dominant strategy (NDS)

Figure 7.1 illustrates the NDS decision tree for the various screening test strategies. NDS is defined as a test strategy in which positive test results are dominated by negative test results, and in which patients with negative screening results receive no further tests. In other words this means that all pregnant women are asked to undertake a follow up diagnostic test after screening positive. For example, SIGN 2001 proposed a two-step universal screening NDS approach. In this approach there is only one TP result following positive results in both test 1 (screening test) and test 2 (diagnostic test) in the disease branch. The probabilities along each of the TP pathways represent the probabilities of cases detected for each strategy.

In another example, SIGN 2010 proposed the use of selective screening, whereby the decision tree commences with the true risk factors stage. This model considers risk factor screening to be a selective screening tool. Therefore, if a patient is negative for risk factors the decision pathway subsequently terminates; those with disease being classified as FN and those without as TN.

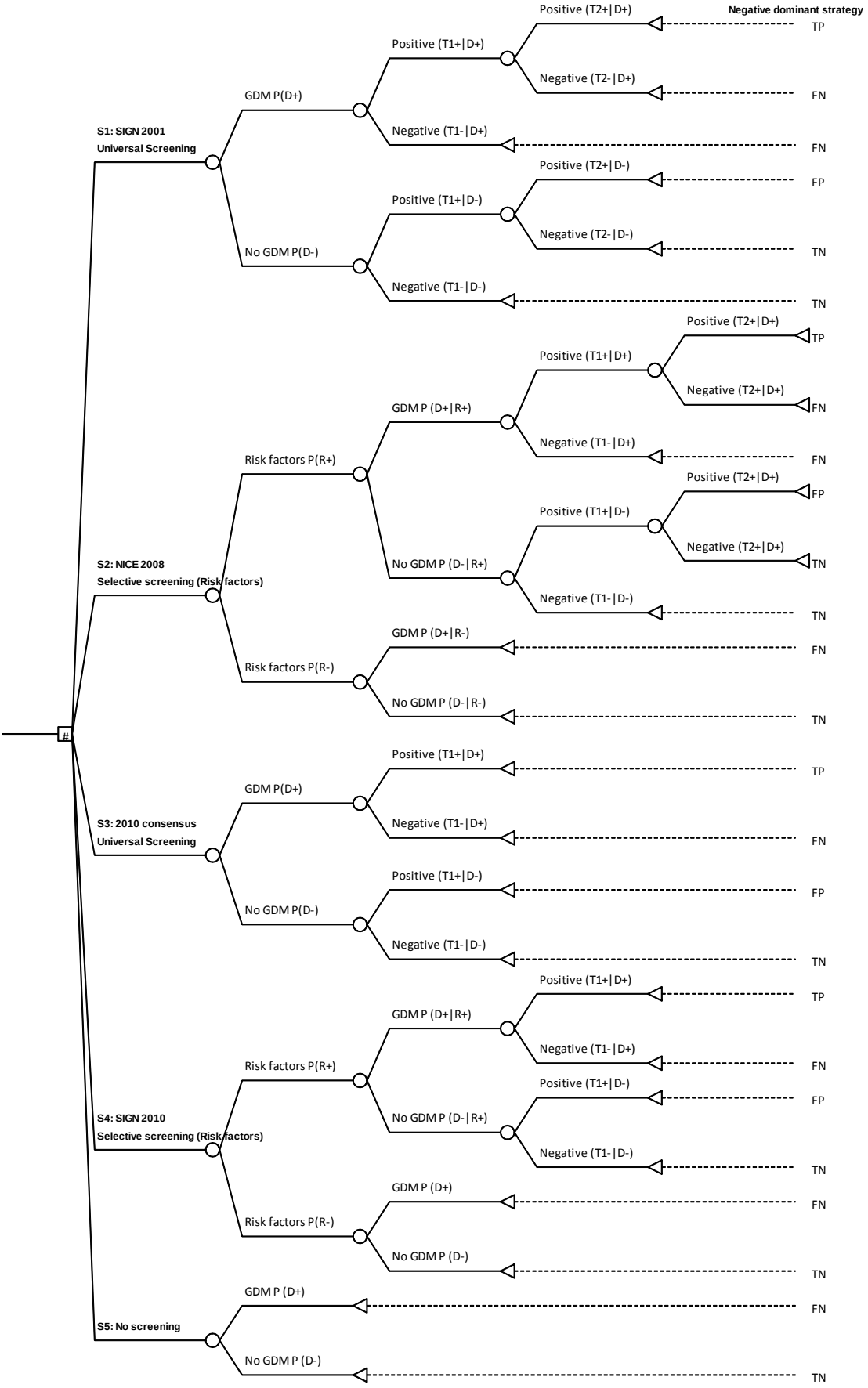


Figure 7.1 Decision tree model for screening tests for GDM (negative dominant strategy)

7.4.2 Case identification: decision tree structure for GDM screening tests in the positive dominant strategy (PDS)

Another test combination approach which needs to be considered for screening tests is the PDS. PDS is defined as a test strategy in which negative test results are dominated by positive test results. In this strategy patients with positive screening results do not receive additional tests, and all patients that have negative screening tests undertake follow up diagnostic tests. Figure 7.2 illustrates the PDS decision tree for the five different strategies. For example adapting SIGN 2001 to a PDS approach, it can be seen that there are two TP results, one following a positive screening test in the disease branch and the second resulting from a negative screening test and positive diagnostic test, in the same disease branch. Therefore, to calculate the probability of cases detected for this strategy, the probabilities for both TP pathways were added together.

Similarly, considering selective screening in terms of PDS, in which risk factors are considered to be a screening test tool, patients that test positive for risk factors receive no further tests and the decision pathway is terminated. In the arm for positive risk factors, patients with disease are classified as TP while mothers without disease are FP.

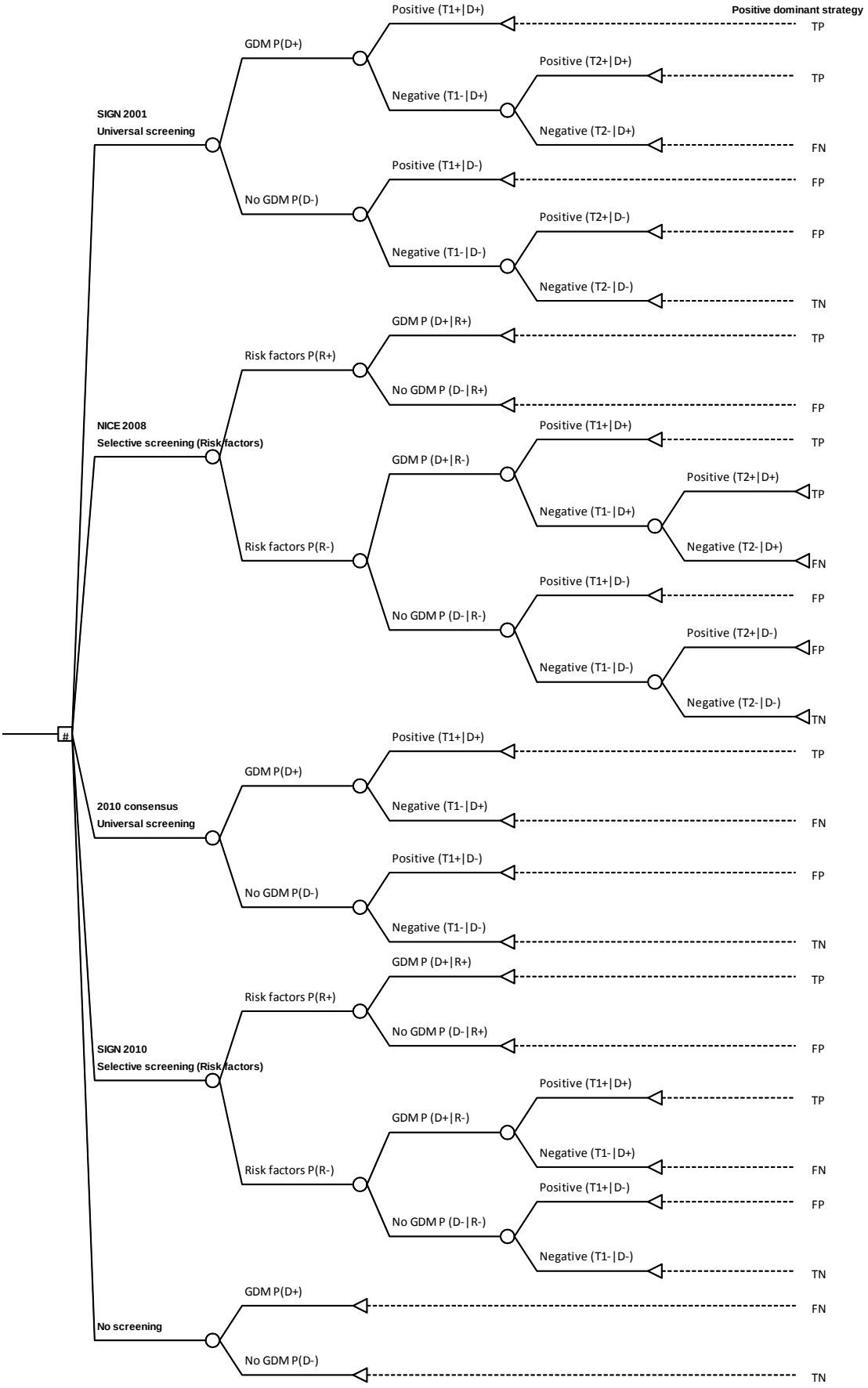


Figure 7.2 Decision tree model for screening for GDM (positive dominant strategy)

7.5 Parameters in economic evaluation modelling for gestational diabetes mellitus: case identification

Having stated and determined the perspective of the model, the appropriate structure of the model was established. Following this it is necessary to identify available data to populate the model, as described in chapter 2, section 2.7.3. The process of populating screening test models involves bringing together all the relevant evidence and synthesizing it appropriately in terms of input parameters within the model. The search terms that were searched from MEDLINE and EMBASE include prevalence of GDM, prevalence of GDM with present or absent risk factors, screening test accuracy and screening costs. The base-case estimates were obtained from various data sources in the published literature. Importantly, parameter estimates were obtained through systematic review and meta-analysis, where not available however, randomised controlled trials and prospective cohorts were used. Expert opinions or internal data from the author's institution were used when no other sources of information were available.

7.5.1 Baseline probabilities in case identification

The essential clinical parameters used to populate the screening test model for GDM include base-case prevalence and screening test accuracy for GDM, both of which were directly converted into probabilities. If multiple data is available, data can be pooled from trials to calculate the baseline risk parameter, as is explained in section 8.3.3. The principles for choosing the most relevant evidence for case identification are discussed in this section.

Prevalence

The construction of the decision tree for the economic evaluation model for GDM screening tests begins with true disease status in the universal screening approach. The primary branch probabilities indicate the probability of outcomes for the patient within a specific population, in terms of presence and absence of disease in universal screening. Prevalence studies of GDM have been published in the past couple of decades (NICE, 2008d) (Jimenez-Moleon et al., 2002). However, in this model, prevalence rates are considered from articles that studied GDM from the perspective of establishing prevalence. Papers that presented results for prevalence but where studied from the perspective of establishing the accuracy of screening were omitted.

The NICE study reports the estimated incidence of gestational diabetes as 3.5% (NICE, 2008d) whilst in the classical prevalence study it ranged from 1 – 16 % (King, 1998). In the systematic review of population-based studies from chapter 5, prevalence was presented as between 1.35% and 12.80%, as shown in Table 7.4. In summary, due to the absence of prevalence estimation in Scotland and following expert advice (Robert Lindsay), the prevalence for GDM of 3.5% reported by NICE is used to represent the baseline parameter for P (D+) in this model. The NICE report was chosen because its data is based on that from England & Wales, which has a similar population demographic to Scotland (The Scottish Government, 2011) (Office for National Statistics, 2012).

Table 7.4 List of baselines and range of relevance parameters with references (disease status)

Variables	Base case	Range	Details	Data source
Prevalence				
GDM in pregnancy (D+)	3.5	NA	Report by NICE	(NICE, 2008d)
	3.3	2.57-4.21	Cross-sectional in Spain	(Jimenez-Moleon et al., 2002)
	3.1	1.35-7.60	Systematic review	Chapter 5
Positive risk factor given that GDM positive $(R+\mid D+)$	38.0	NA	Study in Philippines	(Lim-Uy et al., 2010)
	68.2	NA	Retrospective in Turkey	(Karcaaltincaba et al., 2011)
Positive risk factor given that GDM negative $(R+\mid D-)$	39.2	NA	Observational in Australia	(Moses et al., 1995)
	37.3	NA	Cross-sectional in Thailand	(Sumeksri et al., 2006)
Positive GDM given that risk factor positive $(D+\mid R+)$	50.0	NA	Systematic review	(Scott et al., 2002)
	44.0	NA	Report by NICE	(NICE, 2008d)

For the selective screening protocol, the decision tree begins with risk factors, for which prevalence values are required; denoted by P(R+) and P(R-) in the decision tree model. The branch following risk factors is for true disease status denoted by P $(D+\mid R+)$, P $(D-\mid R+)$, P $(D+\mid R-)$ and P $(D-\mid R-)$. For example, P $(D+\mid R+)$ represents the probability of being positive for GDM given that risk factors are present. This study used Bayes' Rule for calculating conditional probabilities for all of the base line prevalence's (Lesaffre et al., 2007). Conditional probability is a very helpful technique and is outlined in section 3.3. To calculate conditional probabilities for all four of those conditions three values are required; the base line prevalence of GDM P(D+), the probability of having risk factors given that a mother has GDM P $(R+\mid D+)$ and the likelihood of having risk factors given that a mother does not have GDM P $(R+\mid D-)$. The previous section produced a base line prevalence of GDM P (D+) of 3.5%. However, two studies have reported the baseline of P $(R+\mid D+)$ and pooling the values in the articles

produces a figure of 47.2% with a range of 38.0% to 68.2% (Lim-Uy et al., 2010) (Karcaaltincaba et al., 2011). Similarly, two articles have reported the baseline of P $(R+|D-)$ and by pooling these data, an average of 37.8% is calculated with a range of 36.2% to 50.0% (Moses et al., 1995) (Sumeksri et al., 2006). From the above figures, the conditional probability can be calculated, as shown below.

The above information provides the following values and as shown in Figure 7.3.

$P(D+) = 0.035$

$P(R+|D+) = 0.472$

$P(R+|D-) = 0.378$

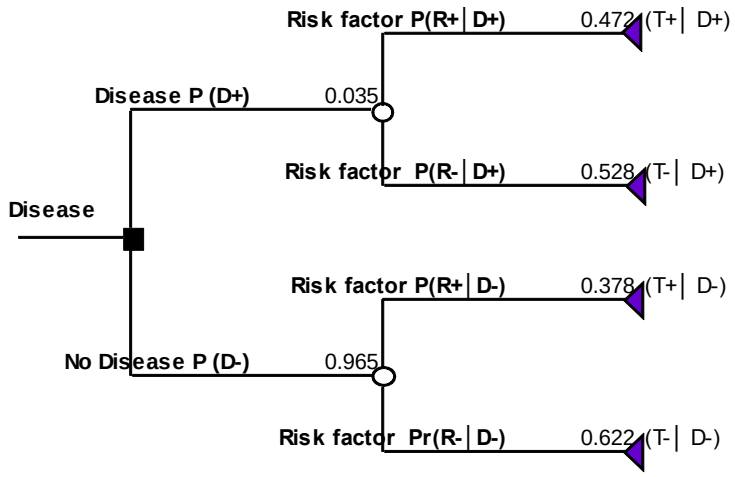


Figure 7.3 Decision tree for risk factors screening

Given this information, we can compute the posterior probability of having GDM given that a mother presents with risk factors.

$$P(D+|R+) = \frac{P(R+|D+)P(D+)}{P(R+|D+)P(D+)+P(R+|D-)P(D-)}$$

$$= \frac{(0.472*0.035)}{(0.472*0.035)+(0.378*0.965)}$$

$$= 0.043$$

We can also compute the posterior probability of not having GDM given the absence of risk factors.

$$P(D-|R-) = \frac{P(R-|D-)P(D-)}{P(R-|D-)P(D-)+P(R-|D+)P(D+)}$$

$$= \frac{(0.622*0.965)}{(0..622*0.965)+(0.528*0.035)}$$

$$= 0.970$$

There is therefore a high probability of 97% for GDM when there is an absence of risk factors. However for the estimated probability of GDM in the presence of positive risk factors, the result is only 4.3%. The lower number of $(D+|R+)$ was a direct result of the prevalence figure of 3.5% that was used to calculate the conditional probability of risk factors based on the prevalence of GDM in a low risk factor population.

These results can be applied to the decision tree to calculate the probability of P(R+) and P(R-) for GDM, as show in Figure 7.4. Using the same information, the baseline with and without risk factors of GDM can be calculated.

$$P(R+) = P(D+)P(R+|D+)+P(D-)P(R+|D-)$$

$$= (0.035*0.472) + (0.965*0.378)$$

$$= 0.381$$

$$P(R-) = P(D+)P(R+|D+)+P(D-)P(R-|D-)$$

$$= (0.035*0.528) + (0.965*0.622)$$

$$= 0.619$$

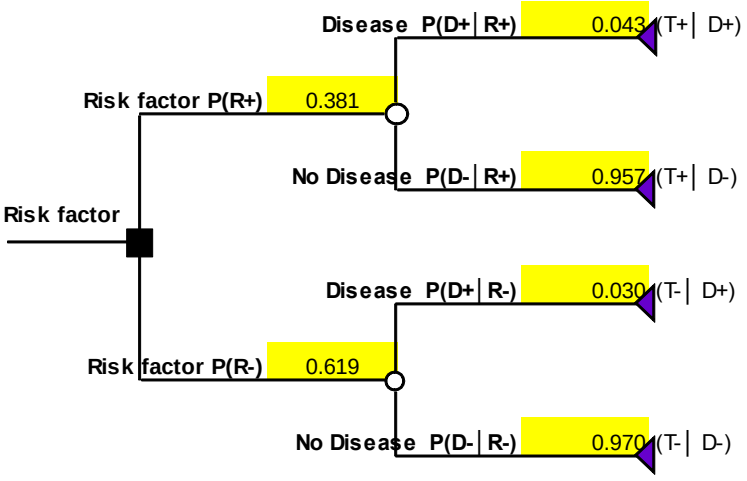


Figure 7.4 Decision tree for calculating the posterior probability

Applying the Bayes' Rule, the baseline risk factor probability P(R+) of GDM was calculated as 38.1%, whereas without risk factors P(R-) it was found to be 61.9%.These numbers were then used in the GDM decision tree for selective screening. In comparison, NICE reported a 44% prevalence for risk factors, which included BMI >30, a high risk ethnic group, previous gestational diabetes and a family history of diabetes (NICE, 2008d), as shown in Table 7.4. Likewise, a systematic study indicated that about 50% of women had historical and clinical risk factors for GDM (Scott et al., 2002). The maximum and minimum values for prevalence observed in these articles ranged from 38.1% to 68.2%.

For the probabilistic analysis, Beta distribution is appropriate for the prevalence data. Parameter estimates were derived from a systematic review of the prevalence of GDM. A summary of the disease status and probabilistic distribution are shown in Table 7.5.

Table 7.5 Summary of clinical parameters in the model (disease status)

Variables	Parameter estimated	Distribution	Parameters of distribution		Data source
			Alpha	Beta	
GDM P(D+)	0.035	Beta	44	1,219	(NICE, 2008d) (Jimenez-Moleon et al., 2002)
Women with risk factor P(R+)	0.381	Beta	481	782	(Moses et al., 1995) (Sumeksri et al., 2006)
Positive for GDM given that risk factors positive P $(D+\mid R+)$	0.043	Beta	20	462	Baye's Theorem
Negative for GDM given that risk factors negative P $(D-\mid R-)$	0.970	Beta	24	757	Baye's Theorem

Screening and diagnostic test accuracy for GDM

The decision tree begins with disease status, and subsequent branching represents possible test outcomes. These pathways indicate the accuracy of the test results as mentioned in chapter 3 section 3.3.3. The pathways that represent test outcomes have sensitivity and specificity values outlined. The most important point of a screening and diagnostic test is to use it to make a diagnosis. This study includes five different screening strategies and uses various screening test techniques such as RPG, FPG, and 75g OGTT. Numerous published articles reported sensitivity and specificity in various screening test techniques. A systematic review in accuracy of RPG reported sensitivity and specificity within a 95% confidence interval (CI) (van Leeuwen et al.,

2010). This model uses those two published papers as the baseline case of RPG. For RPG, sensitivity was reported as 64.00 (95% CI 25.00-75.00), with a corresponding specificity of 80.00 (95% CI 37.00-97.00).

Three studies of the accuracy of screening with FPG reported various values of sensitivity and specificity with a 95% CI (Reichelt et al., 1998) (Senanayake et al., 2006) (Riskin-Mashiah et al., 2010). From these three papers the mean baseline of sensitivity was calculated to be 78.1 with a range of 12.00 and 97.30, and likewise the mean baseline of specificity was found to be 69.57 with a range of 28.90 to 99.50.

In all studies, the accuracy of screening tests for various screening test techniques use 75g OGTT and 100g OGTT as a gold standard to diagnose GDM. However as all four strategies adopted by this model use the 75g OGTT as a diagnostic test, this study assumes 75g OGTT as a gold standard with a sensitivity and specificity of 100%. In addition, there is no trade-off between sensitivity and specificity, and are treated independently. The list of test accuracies is shown in Table 7.6.

Table 7.6 List of baselines and range of relevance parameters with references (test accuracy)

Variables	Base case	Range	Details	Data source
RPG sensitivity	64.00	25.00-75.00	Systematic review	(van Leeuwen et al., 2010)
RPG specificity	80.00	37.00-97.00	Systematic review	(van Leeuwen et al., 2010)
FPG sensitivity	90.00	88.00-94.00	Various threshold ranges	(Reichelt et al., 1998)
	61.50	15.60-80.00	Various threshold ranges	(Riskin-Mashiah et al., 2010)
	82.70	12.00-97.30	Various threshold ranges	(Senanayake et al., 2006)
FPG specificity	66.00	51.00-78.00	Various threshold ranges	(Reichelt et al., 1998)
	75.70	46.90-95.90	Various threshold ranges	(Riskin-Mashiah et al., 2010)
	67.00	28.90-99.50	Various threshold ranges	(Senanayake et al., 2006)
75g OGTT sensitivity	100.00		Reference test	
75g OGTT specificity	100.00		Reference test	

RPG=Random plasma glucose, FPG=Fasting plasma glucose, 75g OGTT= 75g Oral glucose tolerance test

The test accuracy and distribution used in the probabilistic analysis are reported in Table 7.7. Sensitivity and specificity are both probability parameters and the data is binomial. Beta distributions have therefore been used in the probabilistic analysis. All the above papers on test accuracy reported mean values with associated confidence intervals (CI) for both sensitivity and specificity. Consequently, to fit with a beta distribution the Method of Moments was employed by using the sensitivity and specificity estimates and the respective standard errors calculated using the upper and lower confidence limits as mentioned in 2.15.3 (Briggs et al., 2006). There is no uncertainty for

75g OGTT, everyone diagnosed with GDM will undergo this test, and therefore this was held constant in the probabilistic analysis.

Table 7.7 Screening and diagnostic test accuracy parameter estimates

Test Accuracy	Parameters estimate	Distribution	Parameters of distribution		Data source
			Alpha	Beta	
RPG sensitivity	0.640	Beta	8.423	4.748	(van Leeuwen et al., 2010)
RPG specificity	0.800	Beta	4.664	1.166	(van Leeuwen et al., 2010)
FPG sensitivity	0.781	Beta	2.042	0.574	(Reichelt et al., 1998) (Senanayake et al., 2006) (Riskin-Mashiah et al., 2010)
FPG specificity	0.696	Beta	3.845	1.682	(Reichelt et al., 1998) (Senanayake et al., 2006) (Riskin-Mashiah et al., 2010)
75g OGTT sensitivity	1.000	Deterministic			Reference test
75g OGTT specificity	1.000	Deterministic			Reference test

7.5.2 Cost estimates for case identification

All relevant costs and consequences of outcomes should be related to the decision making perspective. The health economic analysis performed in this model is performed from the health care provider perspective, where all relevant screening test costs tend to focus on direct costs only. However, the productivity costs of screening tests include inventory carrying costs; the time taken to compound or dispense drugs in pharmacies; the time nurses take to administer drugs and even the allocation of hospital overhead costs. Likewise productivity costs include the costs of screening for GDM, such as the costs of materials used in various screening and diagnostic tests as well as staff time in performing the tests and biochemical analysis. The costs of various screening tests, derived from published literature, are presented in this section. In order to represent the costs in pounds sterling (£) and at 2011 prices, all relevant costs in different currencies were converted into pounds sterling at the same rate for each year. Next, the Hospital and Community Health Services (HCHS) pay and inflation report was used to adjust costs in other years to those in 2011 (PSSRU, 2010). Prices presented in the report were multiplied by annual percentage increases to calculate expected prices in 2011.

In this model, screening costs include the cost of initial screening test programmes (risk factor screening, RPG and FPG), and the cost of diagnostic tests performed to confirm the results for pregnant women who screen positive with 75g OGTT. The costs of screening tests were retrieved from various published literature. For example, a costing report for antenatal care incorporating diabetes in pregnancy was produced by NICE to provide an implementation tool to estimate the financial impact on

the National Health Service (NHS) (NICE, 2008a). All costs and activity assessments in the report were estimated based on a number of assumptions. The costs of nursing and midwifery assistant time and the costs of biochemical tests were included. Other costs were retrieved from a systematic review of economic evaluations of screening tests for GDM (Scott et al., 2002). This review includes the costs of materials used in the various screening and diagnostic tests, staff time in taking the tests, and biochemical analysis. Another published paper, in Italy, presented intensive metabolic management of GDM, which was obtained by reimbursements for hospitalisation based on disease related groups (DRG) (Di Cianni et al., 2002). In addition, with risk factor screening tests, it is assumed that all pregnant women who receive selective screening (risk factors) do so during their initial visit to the clinic. The unit cost of risk factor screening in the NICE report is calculated using nursing costs based on an average time of five minutes spent at a clinic, as shown in Table 7.8.

Table 7.8 The list of the resources' cost of screening test

Variables	Base case	Range	Details	Data source
RPG/FPG	4.84	NA	Price yr 2002*	(Scott et al., 2002)
	3.60	NA	Price yr 2008**	(NICE, 2008a)
	11.60	NA	Price yr 2002***	(Di Cianni et al., 2002)
75 g OGTT	23.36	NA	Price yr 2002*	(Scott et al., 2002)
	18.76	NA	Price yr 2008**	(NICE, 2008a)
Selective screening	4.95	NA	Price yr 2002*	(Scott et al., 2002)
	9.50	NA	Price yr 2008**	(NICE, 2008a)

* Cost data generated by Scottish Health Purchasing Information centre (SHPIC)/Systematic review for screening test and economic evaluation

** Costing report based on a national tariff

*** Reimbursements for hospitalisation are based on DRG

The cost estimates and distributions used in case identifications are reported in Table 7.9. In probabilistic analysis, normal distributions are considered suitable to represent the cost parameters, however, the characteristic costs of screening and diagnostic tests are the charge price and fixed cost. Therefore, this model presents screening and diagnostic test costs as deterministic values, as NICE have established the price of its tests. In risk factor screening, the unit cost is not the charge price, therefore this model assumes the unit cost of risk factor screening to have a normal distribution. Unfortunately, there was no information available on the variance of mean for unit costs, so the coefficient of variation was used in order to relate the variance to the mean value. Although 0.4 is an arbitrary value, it was chosen to give a large variance (Briggs, 2007).

Table 7.9 Costs for Screening test parameter model

Variables	Parameter estimate	Distribution	Parameters of distribution	Data source
RPG / FPG	3.60	Deterministic	-	(Scott et al., 2002) (NICE, 2008a)
75g OGTT	18.76	Deterministic	-	(Scott et al., 2002) (NICE, 2008a)
Risk factor screening	9.50	Normal	SE = 0.4	(NICE, 2008a)

7.6 Populating the screening test decision tree for GDM

In the beginning of this chapter, the decision tree was outlined and shown to be constructed of decision and chance nodes pertaining to various screening and diagnostic methods based on various test strategies. Moreover, all the relevant evidence presented in previous sections was used to populate the decision tree for case identification. The relevant parameters were converted into probabilities and expressed as a number between zero and one. Probability can be calculated by dividing the desired outcome by a hundred or by the total number of outcomes. All probabilities can then be applied to the branches of the decision tree to represent the possible events patients may experience at any point in the tree. For example, disease status, the first chance node relates to whether or not a mother experiences disease. The probability of GDM prevalence is applied to the disease positive branch (D+), whereas, 1 minus the probability of GDM prevalence is applied to the disease negative branch (D-). Moving from left to right, all subsequent branches show the probability of sensitivity and specificity for all screening and diagnostic tests. This section presents pathway probabilities and the costs of case identification for GDM, with each strategy being summarised as four possible outcomes of case identification.

7.6.1 Expected case identifications

In the decision tree for GDM, the statistics of disease status and diagnostic outcomes indicate the probability along each of the pathways for prevalence of disease (D+) or without disease (D-), prevalence of risk factors (R+) or without risk factors (R-), test positive (T+) and test negative (T-). The combination of branches in the tree that represent screening for GDM outline a series of pathways along which the mother can pass through. The probabilities show the likelihood of a particular event occurring at branches throughout the GDM decision tree. The screening test strategies employed define pathways in the decision tree, for instance S1: 6 pathways, S2: 8 pathways, S3: 4 pathways, S4: 6 pathways and S5: 2 pathways. The expected outcomes are then

calculated by multiplying the pathway probabilities, which are joint probabilities. The probabilities for each strategy must total 1. For example, in strategy 2 (S2), there are four relevant probabilities, namely the probability of having a positive risk factor (R+), the probability of having GDM given the presence of risk factors $P(D+|R+)$, the probability of being test 1 positive given a positive disease state $(T_1+|D+)$ and the probability of being test 2 positive given a positive disease state $(T_2+|D+)$, all of which have the values 0.381, 0.016, 0.781 and 1.000, respectively. The pathway probabilities are then multiplied along the pathway giving a result of 0.005.

Expected case identifications for NDS

In the negative dominant strategy, a true positive result (cases detected) is taken into account for women who have positive results for both screening tests and diagnosis. Consequently, probabilities in the decision tree pathways for each strategy only present one TP result each, as shown in Table 7.10. More than one FN and TN are likely to be found in an NDS. As a result, the probabilities for FN and TN each have to be presented as one probability by adding the individual values. For example, S1 (NDS) has two FN results, namely 0.000 and 0.013. By adding these FN results together the probability of FN in S1 (NDS) is 0.013. Similarly, two results for TN are found in S1 (NDS) giving a result of 0.965 when added together (0.193 + 0.772).

Combining the test results, assuming diagnostic tests have 100% sensitivity and specificity (gold standard), the probabilities for TN and FN are the same in each strategy, whereas, the probabilities of TP and FN outcomes vary depending on the sensitivity of the test employed. In NDS, S3 (NDS) presents higher case identifications than the other strategies. S3 (NDS) is a one-step approach, which involves screening all pregnancies by a diagnostic test with 100% sensitivity and specificity. Therefore, diagnostic tests with 100% accuracy detect GDM from pregnancy close to the actual prevalence in the population (3.5%).

Table 7.10 Expected case identifications for NDS

Probabiltiy from decision tree pathways			Probability for strategy		
	Disease state	**Probability**		**Disease state**	**Probability**
S1 (NDS)	TP	0.022	**S1 (NDS)**	TP	0.022
	FN	0.000		FN	0.013
	FN	0.013		FP	0.000
	FP	0.000		TN	0.965
	TN	0.193			
	TN	0.772			
		1.000			1.000
S2 (NDS)	TP	0.012	**S2 (NDS)**	TP	0.012
	FN	0.000		FN	0.022
	FN	0.003		FP	0.000
	FP	0.000		TN	0.965
	TN	0.111			
	TN	0.254			
	FN	0.019			
	TN	0.600			
		1.000			1.000
S3 (NDS)	TP	0.035	**S3 (NDS)**	TP	0.035
	FN	0.000		FN	0.000
	FP	0.000		FP	0.000
	TN	0.965		TN	0.965
		1.000			1.000
S4 (NDS)	TP	0.016	**S4 (NDS)**	TP	0.016
	FN	0.000		FN	0.019
	FP	0.000		FP	0.000
	TN	0.365		TN	0.965
	FN	0.019			
	TN	0.600			
		1.000			1.000

Expected case identifications for PDS

Table 7.11 presents the results for expected case identifications in the PDS. There is more than one TP result in each strategy. Therefore, to calculate case detection of each strategy, the probabilities of the TP results in each parthway were added up. For example, two results for PT are found in S1 (PDS) giving a result of 0.035 when added together (0.022+0.013). Patients that have an absence of disease but FP are more likely to be found in the PDS. For example, S1 (NDS) has two FP results, namely 0.193 and 0.000. By adding these FN results together the probability of FP in S1 (NDS) is 0.193. Hence, the case identification rates in all PDS strategies are close to the prevalence in the population (3.5%).

In combination of the test results, assuming diagnostic tests have 100% sensitivity and specificity (gold standard), the probability of TP outcomes in PDS is equal to the prevalence in the population. However, in PDS, the probabilities for TP and FN are the same in each strategy, whereas, the probabilities of FP and TN outcomes vary depending on the specificity of the test employed.

Table 7.11 Expected case identifications for PDS

Probabiltiy from decision tree pathways			Probability for strategy		
	Disease state	Probability		Disease state	Probability
S1 (PDS)	TP	0.022	**S1 (PDS)**	TP	0.035
	TP	0.013		FN	0.000
	FN	0.000		FP	0.193
	FP	0.193		TN	0.772
	FP	0.000			
	TN	0.772			
		1.000			1.000
S2 (PDS)	TP	0.016	**S2 (PDS)**	TP	0.035
	FP	0.365		FN	0.000
	TP	0.015		FP	0.548
	TP	0.004		TN	0.417
	FN	0.000			
	FP	0.183			
	FP	0.000			
	TN	0.417			
		1.000			1.000
S3 (PDS)	TP	0.035	**S3 (PDS)**	TP	0.035
	FN	0.000		FN	0.000
	FP	0.000		FP	0.000
	TN	0.965		TN	0.965
		1.000			1.000
S4 (PDS)	TP	0.016	**S4 (PDS)**	TP	0.035
	FP	0.365		FN	0.000
	TP	0.019		FP	0.365
	FN	0.000		TN	0.600
	FP	0.000			
	TN	0.600			
		1.000			1.000

7.6.2 Expected costs

Each pathway in the GDM tree has associated costs, including risk factor screening, screening and diagnostic tests, depending on the strategy. For example, in the first pathway of S2 (NDS), the significant costs are risk factor screening £9.50, screening tests £3.60 and diagnostic tests £18.76, totalling £31.86. The same principle is used for other pathways in the tree and it can be seen that for some pathways the total sum of the costs remain the same. The expected costs for each strategy can be calculated by

summing the values for the pathway costs weighted by the pathway probabilities. The total costs of risk factor screening, screening and diagnostic tests are shown as the expected value of each pathway. The pathway probabilities represent the sum of the different probabilities for each branch in the pathway. Of those, S2 (NDS) for example, has a pathway probability 0.012 and a pathway value of £31.86, a weighted value can be calculated to give the expected costs of this pathway, equaling 0.39, as shown in Table 7.12. The same principle is used for the other pathways in the tree to calculate expected costs. Adding up the expected costs along the pathways for each strategy can generate expected costs for the strategies. With the same example S2 (NDS), the expected cost is £13.19. Another way of working out the expected costs and consequences for a given option in a decision tree is by rolling back the tree. By doing so, the expected outcomes will be exactly the same as summing across all the pathways.

7.7 Cost per case identification

If a screening test is adopted as standard, other tests cannot then be employed. Considering this problem, a decision has to be made based on the cost and effectiveness of all relevant screening tests. This section discusses incremental cost-effectiveness ratios (ICER). There are four different screening test strategies being considered along with an alternative of 'doing nothing'. The outcomes were measured for cases detected, and defined by the probability of the true positive result in each strategy. Costs were measured in pounds sterling, and were calculated by the summation of all the costs within each strategy.

7.7.1 Cost per case identification in the NDS

For the negative dominant strategy, S1 (NDS) shows a lower expected cost of about £7.64, as shown in Table 7.12. On the other hand, S3 (NDS) is seen to be more effective in detecting cases of GDM with a TP of 0.035. S3 (NDS) is a universal screening test, which involves screening all pregnancies by a diagnostic test with 100% sensitivity and specificity. Diagnostic tests with 100% accuracy detect GDM from pregnancy close to the actual prevalence in the population (3.5%), with expected cost of £18.76. S2 (NDS) has the highest screening cost £31.86, because it includes three different screening methods, namely risk factor screening, screening tests and diagnostic tests.

Table 7.12 Expected costs of case identification in the negative dominant strategy.

	Detection status	Probability	Screening cost	Expected cost
S1 (NDS)	TP	0.022	22.36	0.50
	FN	0.013	25.96	0.05
	FP	0.000	22.36	0.00
	TN	0.965	25.96	7.09
		1.000		**7.64**
S2 (NDS)	TP	0.012	31.86	0.39
	FN	0.022	54.46	0.23
	FP	0.000	31.86	0.00
	TN	0.965	54.46	12.57
		1.000		**13.19**
S3 (NDS)	TP	0.035	18.76	0.65
	FN	0.000	18.76	0.00
	FP	0.000	18.76	0.00
	TN	0.965	18.76	18.11
		1.000		**18.76**
S4 (NDS)	TP	0.016	28.26	0.45
	FN	0.019	37.76	0.18
	FP	0.000	28.26	0.00
	TN	0.965	37.76	16.02
		1.000		**16.65**
S5	TP			
	FN			
	FP			
	TN			
		0		**0.00**

7.7.2 Cost per case identification in the PDS

In the positive dominant strategy, the expected costs of screening in S4 (PDS) are high, about £21.12 per strategy, because it includes two different screening methods, namely risk factor screening and diagnostic tests. S1 (PDS), on the other hand, shows a lower expected cost of about £18.32, as shown in Table 7.13. The expected costs of screening in S3 (PDS) are £18.76, which is a similar value to the expected cost of screening in S3 (NDS).

Table 7.13 Expected costs of case identification in positive dominant strategy.

	Detection status	Probability	Screening cost	Expected cost
S1 (PDS)	TP	0.035	25.96	0.36
	FN	0.000	22.36	0.00
	FP	0.193	25.96	0.69
	TN	0.772	22	17.27
		1.000		**18.32**
S2 (PDS)	TP	0.035	54.46	0.48
	FN	0.000	31.86	0.00
	FP	0.548	54.46	5.86
	TN	0.417	31.86	13.30
		1.000		**19.64**
S3 (PDS)	TP	0.035	18.76	0.65
	FN	0.000	18.76	0.00
	FP	0.000	18.76	0.00
	TN	0.965	18.76	18.11
		1.000		**18.76**
S4 (PDS)	TP	0.035	37.76	0.69
	FN	0.000	28.26	0.00
	FP	0.365	37.76	3.47
	TN	0.600	28.26	16.96
		1.000		**21.12**
S5	TP			
	FN			
	FP			
	TN			
		0		**0.00**

7.7.3 Incremental cost-effectiveness ratios for case identification in GDM screening

The costs per case identification in the above section can be used in a cost-effectiveness analysis where the costs and cases of GDM detected (outcomes) are compared for both NDS and PDS. The 10 strategies (5 NDS and 5 PDS) were used to calculate the incremental cost effectiveness ratio (ICER). Of those, both the NDS and PDS results for Strategy 5 (S5) “do nothing” were grouped and presented as a single strategy comparator. Likewise, the NDS and PDS results for Strategy 3 were grouped as a single strategy because they presented with identical results for both costs and cases detected. So, there were a total number of 8 strategies.

Table 7.14 shows the expected costs and case identification of 8 screening test strategies. The screening test strategies were ranked according to costs. S2 (NDS) and S4 (NDS) are seen to be more costly and with less cases detected than S1 (NDS). Similarly, S3, S2 (PDS) and S4 (PDS) were more costly and had less case detections compared with S1 (PDS).

Table 7.14 Cost and effectiveness in difference screening test strategy

Screening strategies	Costs (£) [C]	Case identification [E]
S5	0.00	0.00000
S1(NDS)	7.64	0.02230
S2(NDS)	13.19	0.01234
S4(NDS)	16.65	0.01580
S1(PDS)	18.32	0.03484
S3	18.76	0.03484
S2(PDS)	19.64	0.03483
S4(PDS)	21.12	0.03483

The CE plane of cases detected was plotted on a graph for each strategy, with an x-axis of probability of cases detected and y-axis of screening costs, as seen in Figure 7.5. For example, for S1 (PDS), cases detected and screening costs were 0.032 and £18.32, respectively. The ICER is the sloping line between screening strategies in Figure 7.5. For example, the sloping blue line between S1 (NDS) and S1 (PDS) has a value of 851.86, as discussed below.

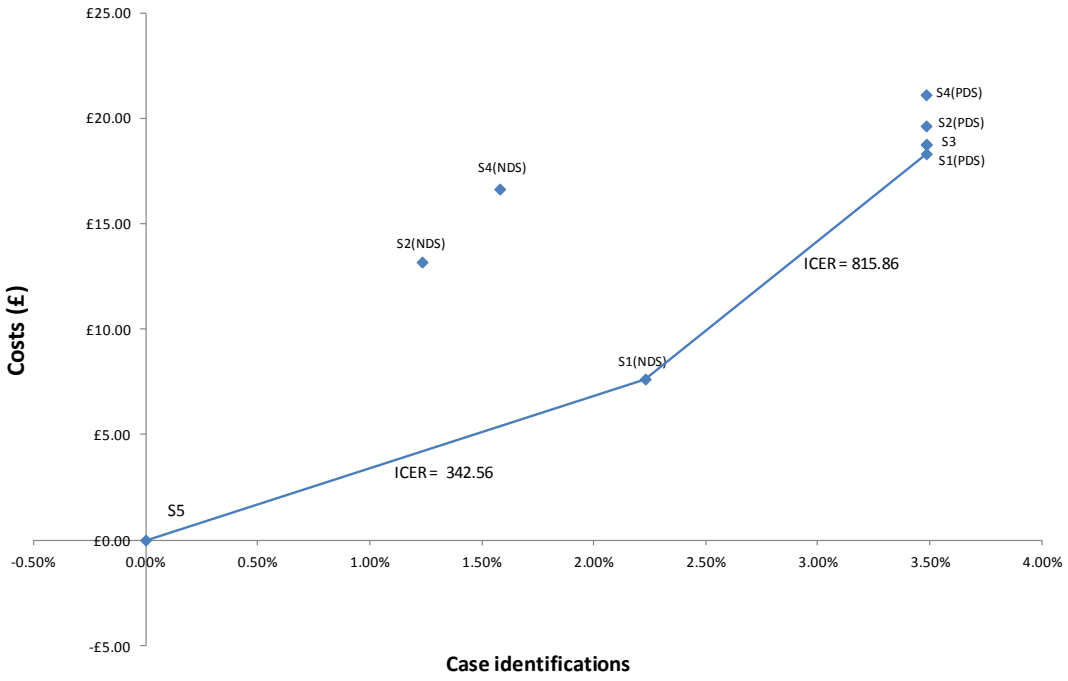


Figure 7.5 the cost-effectiveness plane for cases detected for GDM screening tests

However, it is not possible to draw any definitive conclusions regarding the cost-effectiveness of different screening tests based on the average cost effectiveness ratios (ACER) as shown in the third column of Table 7.14. Because of this, the ICER has to be calculated to summarise the cost per case detected for GDM. To estimate the ICER's, the alternative screening test strategies were ranked according to costs. Each ICER [△C/△E] was then calculated by dividing the incremental cost [△C] by the incremental effect [△E] for each screening test. From these results, S2 (NDS) and S4 (NDS) are seen to be more costly and with less cases detected than S1 (NDS). So, S1 (NDS) is dominant to S2 (NDS) and S4 (NDS). The dominated screening test strategies are then excluded and the ICER recalculated. Similarly, S3, S2 (PDS) and S4 (PDS) were more costly and had less case detections compared with S1 (PDS). Therefore, S3, S2 (PDS) and S4 (PDS) are dominated by S1 (PDS). Having excluded S2 (NDS), S4 (NDS), S2 (PDS) and S4 (PDS), the ICER was recalculated for S5, S1 (NDS) and S1 (PDS), as shown in Table 7.15.

Table 7.15 Incremental cost-effectiveness ratios along the efficacy frontier when exclusion of more costly and less effective alternatives

Screening strategies	Costs (£) [C]	Case identification [E]	Incremental costs [△C]	Incremental effect [△E]	ICER [△C/△E]
S5	0.00	0.00000	0.00	0.000	0.00
S1(NDS)	7.64	0.02230	7.64	0.022	342.56
S1(PDS)	18.32	0.03484	10.68	0.013	851.86

7.7.4 Probabilistic result

Probabilistic sensitivity analyses explore uncertainty in cost-effectiveness outcomes. The economic models of screening tests for GDM are analysed probabilistically using Monte Carlo simulation (1000 iterations) in order to determine expected costs, outcomes and cost-effectiveness. The costs relate to screening test costs in each strategy while outcomes relate to case identifications. The results of the 1000 replications from the model are presented on the cost-effectiveness plane, as shown in Figure 7.6. S3 is shown as a line because the screening test costs in this model are fixed cost and it is a one step approach strategy, with no uncertainty in the costs.

Figure 7.6 Results of 1000 Monte Carlo simulation evaluations for screening tests for GDM on the cost-effectiveness plane

7.7.5 Decision uncertainty

The aim of calculating a CEAC is to summarise and illustrate graphically the probability that a particular intervention or screening test is the optimal choice over a wide range of values for the ceiling ratio (λ). The CEAC in terms of screening tests for GDM shows the decision uncertainty surrounding the cost-effectiveness of each strategy.

The uncertainty in cost-effectiveness estimated for screening for GDM is illustrated in a CEAC in Figure 7.7. The curve presents the probability that a screening test for GDM is cost effective at society's willingness to pay for case identification. The

CEAC presents the results in cost per quality-adjusted life years (QALYs) which is suitable for the decision maker because it is the acceptable threshold used by decision making institutions (Roberts et al., 2012b). To interpret the results in the absence of QALYs, this thesis presents the outcomes of case identifications of GDM in terms of natural units. In the UK, the willingness to pay threshold recommended by the NICE is £20,000 - £30,000 per QALYs (NICE, 2008c). This means that if the ICER for a specific intervention is below the cost effectiveness threshold then society will be willing to fund it, while on the other hand if the ICER is greater than the threshold society will not be willing to fund it. Thus, it is inappropriate to interpret the results using natural units in terms of the arbitrary willingness to pay threshold. However, CEAC illustrates graphically the probability that a particular screening test is the optimal choice over a wide range of values for the ceiling ratio. Therefore, for case identification, a willingness to pay range of £0 - £3,000 is used to present the decision uncertainty surrounding the cost-effectiveness of each screening test strategy. From Figure 7.7, at the willingness to pay of £0 - £3,000, the CEAC illustrates that the probability that S2 (PDS) would be cost-effective is approximately 39%.

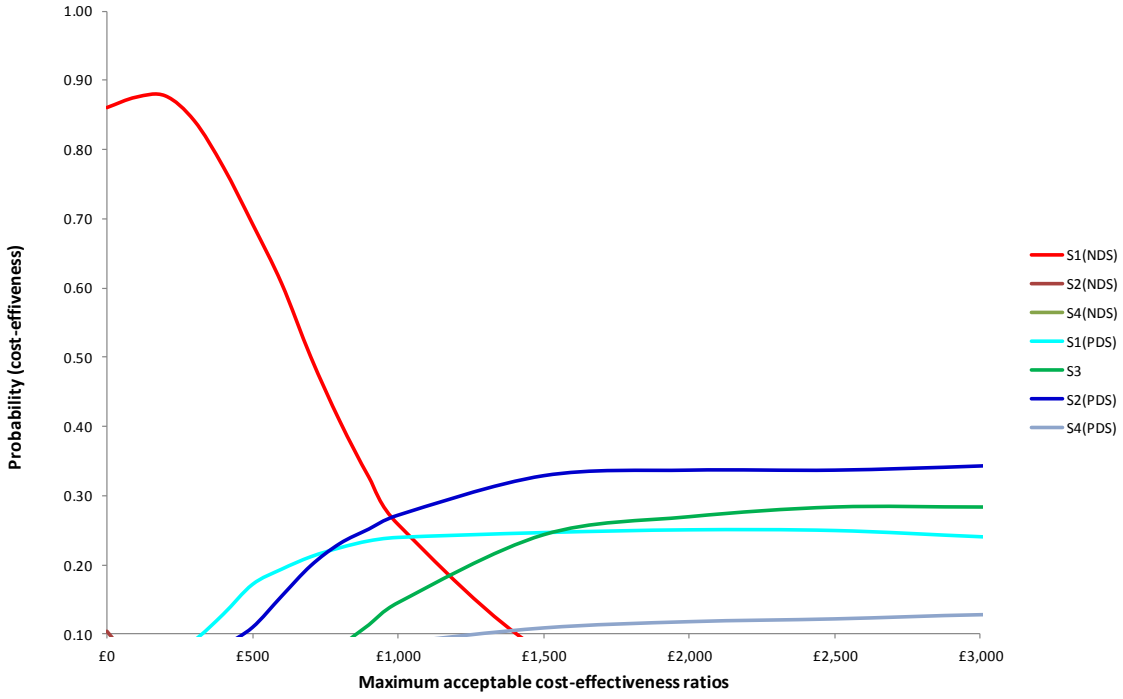


Figure 7.7 Cost-effectiveness acceptability curves of case identifications for GDM

7.8 Discussion

The probabilistic sensitivity analysis suggested that at willingness to pay of £0 to £5,000 the probability of S2 (PDS) being cost-effective is 39%. S2 is a screening test strategy that was published by NICE in 2008 (NICE, 2008b). The NICE guidelines recommend NDS with selective screening that screens all mothers with risk factors at first visit followed by FPG and 75g OGTT. According to the results, S2 using a PDS approach with selective screening followed by a two-step approach using FPG and 75g OGTT yielded the highest rate of case identification and the lowest costs for screening tests for GDM.

No studies carried out a model-based economic evaluation to assess the cost-effectiveness of screening tests for GDM using an outcome of cost per case identifications. In this study, the cost per case identifications ranges from £7.64 - £21.12 depending on the screening test used. However, to interpret the results in the absence of QALY, cost per case identifications of GDM were compared with those in a cost per case screening study in Canada that involved three different screening strategies for GDM, and in which the costs ranged from CA$89.03 – $108.38 (Meltzer et al., 2010). The key difference in cost identification in each of the studies was due the evaluation of direct and productivity costs. The Canadian study included technician time and laboratory analysis as direct costs of the screening tests. While, the productivity costs, which included one-way transportation from home to hospital, the cost of time women spent in one-way transit to the blood test centre and to complete the glucose testing were also accounted for. In contrast, the costs presented in this thesis take into account only the direct costs of the test methods, and are therefore likely to be underestimated. However, the objective of introducing a screening programme is to offer treatment and to prevent short and long term complications of GDM, for this reason case detection is perhaps not an appropriate outcome to use for CEA. The CEA of case identification measures health outcomes in terms of natural units, which do not take into account the impact of FP and FN. Therefore, it is important to analyse the longer term impact of diagnostic tests, and similarly the impact that FP and FN results have on patients. Health related quality of life for short and long term complications may be the appropriate outcome to capture the health outcome impact of FP and FN results. Moreover, it is difficult to interpret the results in the absence of QALYs in CEAC, as mentioned in the previous section.

The combinations of tests in NDS and PDS represent alternative approaches in evaluating the economic assessment of diagnostic tests. Screening test strategies for GDM and screening tests for other diseases usually recommend NDS to screen and

detect disease in individuals in the general population. In NDS, only patients that test positive in the first test (screening test) undergo a second test (diagnostic test), however only a few people test positive in the second test and so only a small proportion of the population with the disease are detected (TP). TN and FN results are therefore the main outcomes of NDS. Therefore there is a high probability that patients without the disease have a negative result in NDS. In NDS, clinicians should be concerned about tests with high sensitivity, as tests that can identify patients with the disease who have positive test results reduce the numbers of FN patients. In PDS, it is patients that are negative in the first test (screening test) that undergo an additional second test (diagnostic test). TP and FP results are the main outcomes of PDS where screening and diagnostic tests involve positive results. In PDS, clinicians should be concerned about tests with high specificity, as tests that can identify patients without the disease who have negative results reduce the numbers of FP patients.

The NDS and PDS approaches allow the clinician to consider test results in terms of the differences in FN and FP test results. The trade-off between sensitivity and specificity or detection and unnecessary testing is at the heart of screening and diagnostic tests and the accuracy of the tests in terms of sensitivity and specificity is an important consideration for clinicians. Even if clinicians are able to keep both specificity and sensitivity high in tests, FP and FN outcomes will still persist.

7.9 Conclusion

An economic model for screening for GDM was been developed and presented as a decision tree model which starts with a decision for the screening strategy followed by true disease status (prevalence), screening and diagnostic test classification (TP, FN, FP, TN) and case detection. The perspective of the model was stated and determined initially, allowing the appropriate structure of the model to be established. Then the necessary available data to populate the model was identified, including the costs for screening tests, prevalence of disease and accuracy of tests.

With regard to the costs per case of GDM identifications, PDS was found to be more costly than NPS. As the population has a low prevalence of GDM (3.5%), the number of cases detected was low and is reflected in the TP result. This chapter evaluated CEA by expressing consequences in terms of natural units, determined here as cases of GDM detected, and the use of which maximises the number of cases identified. However, case detection may not be an appropriate outcome to use for CEA in this study as the objective to introduce a screening programme is to offer treatment and prevent short term and long term complications of GDM. The next chapter will

recalculate CEA by using long and short term complications instead of case identification.

Chapter 8 The decision analysis in gestational diabetes mellitus: treatment and short term and long term complications

8.1 Introduction

This chapter presents the development and results of a decision analysis model of screening tests for GDM including treatment during gestation as well as short term and long term complications. The structure of decision models are based on the nature of the disease and its treatment. The decision analysis for GDM for treatment is discussed at the beginning of this chapter and in the first section the model for treatment during gestation is structured as a decision tree. Following this the parameters and costs relevant to treatment are discussed and selected.

If mothers with GDM are not diagnosed and are treated inappropriately, this can lead to adverse complications in both mother and child. The next section deals with how to construct a decision tree with regard to adverse short term complications. The appropriate parameters and costs relevant to adverse complications are evaluated and discussed. The techniques used to estimate the baseline parameters are also shown. Before calculating the utility for mothers, the duration of pregnancy is required to identify when adverse complications would occur during gestational and postpartum periods. The following section after this briefly discusses the formal methods used to translate a person's perception of quality of life in various health states into Health Related Quality of Life (HRQoL) scores for GDM. Quality adjusted life years (QALY) for short term complications in both mothers and offspring are also evaluated and discussed.

Lastly with respect to long term complications, type 2 diabetes mellitus (DM) can develop after a diagnosis of gestational diabetes in both mothers and offspring. Therefore, life tables and life expectancies are generated and used to calculate lifetime QALY for long term complications in both mothers and offspring with DM. This section also details all relevant information that is used to calculate lifetime QALY. HRQol for type 2 DM in the general population and different age groups are also discussed.

8.2 Decision analysis of screening tests for GDM: treatment

Pregnant women diagnosed with GDM have to be treated to reduce the risks of complications to themselves and their offspring. Two randomised controlled trial (RCT) studies state that the treatment of GDM with nutritional counselling, dietary control and insulin reduced serious adverse complications in both mothers and offspring (Crowther et al., 2005) (Landon et al., 2009). In addition, the Scottish Intercollegiate Guidelines Network for the management of diabetes (SIGN) and the National Institute for Health and Clinical Excellent (NICE) guidelines for antenatal care of diabetes in pregnancy are recommended for the treatment of mothers with GDM with dietary control with glucose monitoring, oral medication and insulin therapy. Management of GDM in these guidelines are the same and similar to the two RCT above, including dietary therapy and blood glucose monitoring on the one hand, and pharmacotherapy (Glibenclamide and Metformin) or insulin therapy (Regular insulin, Insulin aspart and Insulin lispro) on the other. Therefore, treatment of GDM in this study includes dietary control with glucose monitoring, oral medication and insulin therapy based on the SIGN and NICE guidelines for antenatal care of diabetes in pregnancy (SIGN, 2010) (NICE, 2008d). This study assumes that all women identified as having GDM received treatment, that they received this treatment without fail and that all treatment was completed.

8.2.1 Treatment: decision tree structure

The decision trees for the treatment branches are attached after the case identification tree, as shown in Figure 8.1. The four possible outcomes for the identification of GDM include true positive (TP), false negative (FN), false positive (FP), and true negative (TN) results for GDM. Of those outcomes, there are only two possible pathways that take into account mothers that receive treatment. These include a TP outcome, cases where both the disease and the test show positive; and FP outcomes, namely cases without disease but where the test result shows the presence of disease. For the other two case identification results, treatment is not considered and the decision tree terminated at these points. These include FN which are cases where there is disease but the test indicates that there is not; and TP which are cases without disease where the test confirms its absence. This model assumes that pregnant women that have FN and TN results do not receive any treatment until delivery, and are screened once per person.

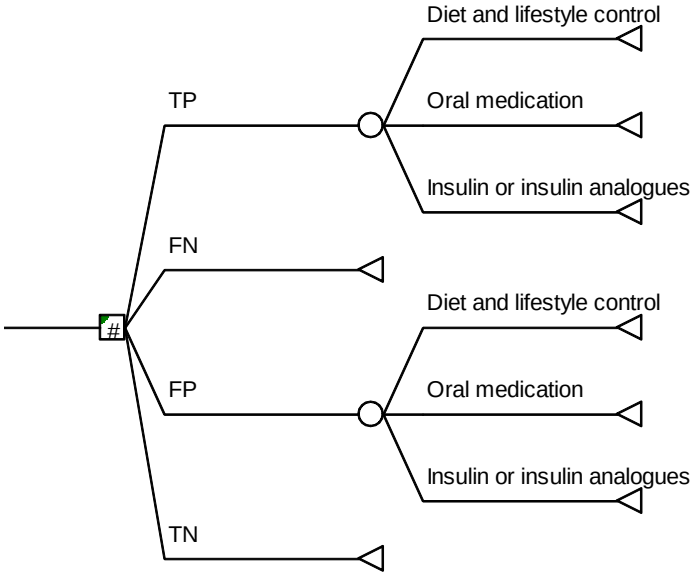


Figure 8.1 Decision tree model for treatment

8.2.2 Treatment: baseline probabilities

Based on treatment guideline form NICE and SIGN, the probability of mother being treated with three different treatment is needed to populated the treatment model. NICE also reported the percentage of treatment for GDM as 65.0%, 20.0%, and 15.0% for diet and lifestyle, oral medication and insulin or insulin analogues, respectively (NICE, 2008a). The list of clinical parameters for treatment and references are shown on Table 8.1.

Table 8.1 List of clinical parameters in treatment.

Variables	Base case	Range	Details	Reference
Diet and life style	65.0	NA	Report by NICE	(NICE, 2008a)
Oral medication	20.0	NA	Report by NICE	(NICE, 2008a)
Insulin or insulin analogues	15.0	NA	Report by NICE	(NICE, 2008a)

This model adopted the NICE figures to calculate and adjust the relevant treatment costs for GDM. This model assumes that all women identified as having GDM receive care management without fail and that their treatment is completed. It is assumed that women with TP and FP outcomes would start treatment at a gestational age of 27 weeks and that this would continue for 90 days. There is no uncertainty in treatment during the gestation period, all treatment during gestation was held constant in this probabilistic analysis, as shown in Table 8.2.

Table 8.2 Clinical parameter of treatment in the model

Variables	Probability	Distribution	Parameters of distribution	Reference
Diet and life style (Blood glucose monitoring)	0.650	Deterministic	-	(NICE, 2008a)
Oral medication	0.200	Deterministic	-	(NICE, 2008a)
Insulin or insulin analogues	0.150	Deterministic	-	(NICE, 2008a)

8.2.3 Treatment: cost estimates

Costs data for treatment were derived from published literature. The cost of implementing each treatment is derived from a health care provider's perspective. All relevant costs derived from different currencies were converted and recalculated using the same technique that was used for the screening tests, as mentioned in section 7.5.2. The main treatment costs are derived from the National Institute for Health and Clinical Excellence report, Antenatal Care, Diabetes in Pregnancy Costing Report, which consists of the cost of oral medication, insulin/ insulin analogues, and blood monitoring (NICE, 2008d). These costs assumed 90 days medication and 12 week monitoring (to allow time for diet and exercise to be tried) during pregnancy. The list of relevant costs and references are shown on Table 8.3.

Table 8.3 List of costs in treatment.

Variable	Base case (£)	Range	Details	Reference
Diet and life style	122.56	NA	Price yr 2008*	(NICE, 2008d)
(Blood glucose monitoring)	146.44	NA	Price yr 2002***	(Moss et al., 2007)
	132.70	NA	Price yr 2002**	(Di Cianni et al., 2002)
Oral medication	3.30	NA	Price yr 2008*	(NICE, 2008d)
Insulin or insulin analogues	63.58	NA	Price yr 2008*	(NICE, 2008d)
	224.90	NA	Price yr 2002***	(Moss et al., 2007)
	120.22	NA	Price yr 2002**	(Di Cianni et al., 2002)

*Costing report based on a national tariff
**Reimbursements for hospitalization are based on DRG
***Hospital cost

All costs used in the decision tree model are summarised in Table 8.4. The treatment costs are taken from the above NHS reports and as they are fixed costs or charge price they are held constant in the probabilistic analysis.

Table 8.4 Costs for Screening test parameter in model

Variable	Point estimate	Distribution	Parameters of distribution	Reference
Diet and life style (Blood glucose monitoring)	122.56	Deterministic	-	(NICE, 2008d)
Oral medication	3.30	Deterministic	-	(NICE, 2008d)
Insulin or insulin analogues	63.58	Deterministic	-	(NICE, 2008d)

8.2.4 Expected treatment costs for GDM

The costs for each type of treatment presented in Table 8.5 are weighted for probability and costs. These weightings can be aggregated to give the total expected treatment costs for GDM, £89.86 per person.

Table 8.5 Treatment costs for GDM

Variable	Probability	Cost (£)	Expected cost (£)
Diet and life style (Blood glucose monitoring)	0.650	122.56	79.66
Oral medication	0.200	3.30	0.66
Insulin or insulin analogues	0.150	63.58	9.54
Total			89.86

As mentioned in section 8.2.1, there are only two possible case identification outcomes which result in treatment being received, namely TP and FP results. The expected cost of treatment for GDM of £89.86 is shown in the treatment arms where women test positive for GDM (TP) and where women test positive in the absence of disease (FP), as shown in Table 8.6. On the other hand, in the non treatment arms, treatment costs per person during pregnancy are invisible. Previous economic evaluations included delivery care costs with treatment during gestational period. This may lead to a double count in adverse complication costs for mothers who require caesarean section. Therefore, this model was designed to calculate normal delivery care costs together with short term complications, as discussed in the next section.

Table 8.6 Treatment costs based on case identification

Variable	Expected cost (£)
True positive (TP)	89.86
False negative (FN)	
False positive (FP)	89.86
True negative (TN)	

8.3 Decision analysis of GDM: short term complications

Diabetes in pregnancy involves a risk of adverse complications to the mother and the developing foetus. Many clinical guidelines recommend the management of diabetes and its complications for women who wish to conceive and those who are already pregnant (NICE, 2008d) (SIGN, 2010). Effective management of GDM, including blood glucose monitoring, oral medication and regulating insulin, reduce adverse effects for both mother and child. GDM is associated with substantial maternal and prenatal complications, as mentioned in section 4.3. Risks for Short term complications (SC) such as preeclampsia, hypertensive disorder, caesarean sections, macrosomia, shoulder dystocia and metabolic problems as mentioned in section 4.4 were selected based on the adverse complications that were commonly used in previous studies of economic evaluation of screening tests for GDM, as outlined in chapter 6. Some studies

included perinatal adverse outcomes, namely maternal and neonatal death or stillbirth in economic evaluations of screening tests for GDM (Lohse et al., 2011). It is reported in randomised clinical trials that such cases of adverse outcomes are very rare (Crowther et al., 2005) (Landon et al., 2009). Thus, this model decided to exclude maternal and neonatal death as well as stillbirth from short term adverse complications. This study assumes that all mothers and offspring have the possibility of developing all of the adverse complications included in the study.

This model estimates the cost-effectiveness of screening for and treating GDM, considering short term complications in both mothers and offspring for four alternative strategies versus a strategy of “no screening”. In the UK, the average age at first pregnancy is 26 years old (Office for National Statistics, 2013). This model therefore assumes all pregnant women experience their first pregnancy at age 26. The model for the utility for mothers over a 1 year period after delivery was considered from age 27. Therefore, the calculations for utility over a 1 year period after delivery were performed for mothers up until age 28. Costs and outcomes are considered for both mothers (pregnancy period plus 3 months after delivery) and offspring (neonatal period plus 11 months of infancy) over a period of one year. However, a discount rate was not applied to this analysis, as is the case with economic evaluations with short time horizons (1 year).

8.3.1 Short term complications in mothers: decision tree structure

The SC decision trees for both mothers and offspring are attached to the decision tree for GDM screening tests. They are attached after the treatment tree for mothers who get treatment during gestation, both TP and FP. Whereas, another point where SC decision trees attach is after case detection in mothers who are TN, and FN, and who are also able to present with adverse complication. The following section illustrates two different tree maps for both mothers and offspring.

In the economic evaluations performed in previous GDM studies, the SC decision trees were constructed in either the gestation or post partum periods after pregnancy. To construct the decision tree in only one period of gestation might not represent the actual period within which SC occurs. Therefore, the decision tree for SC in this model reflects both periods after pregnancy, i.e. the gestation and postpartum periods. Preeclampsia and hypertension occur during the gestation period. Caesareans section and normal delivery is a technique for childbirth which may affect the mother

during the postpartum period. Therefore, the decision tree for SC in mothers starts with the chance node of three main outcomes including, preeclampsia hypertension, and no AC which represent the gestation period. Each of those outcomes is then followed by chance nodes for whether or not a mother experiences caesarean section or normal delivery. From these branches, the probability of the event is represented. Moreover, normal delivery is assumed as natural childbirth without the use of any medical interventions and without any complication during postpartum. Figure 8.2 illustrates the decision tree of SC in mothers.

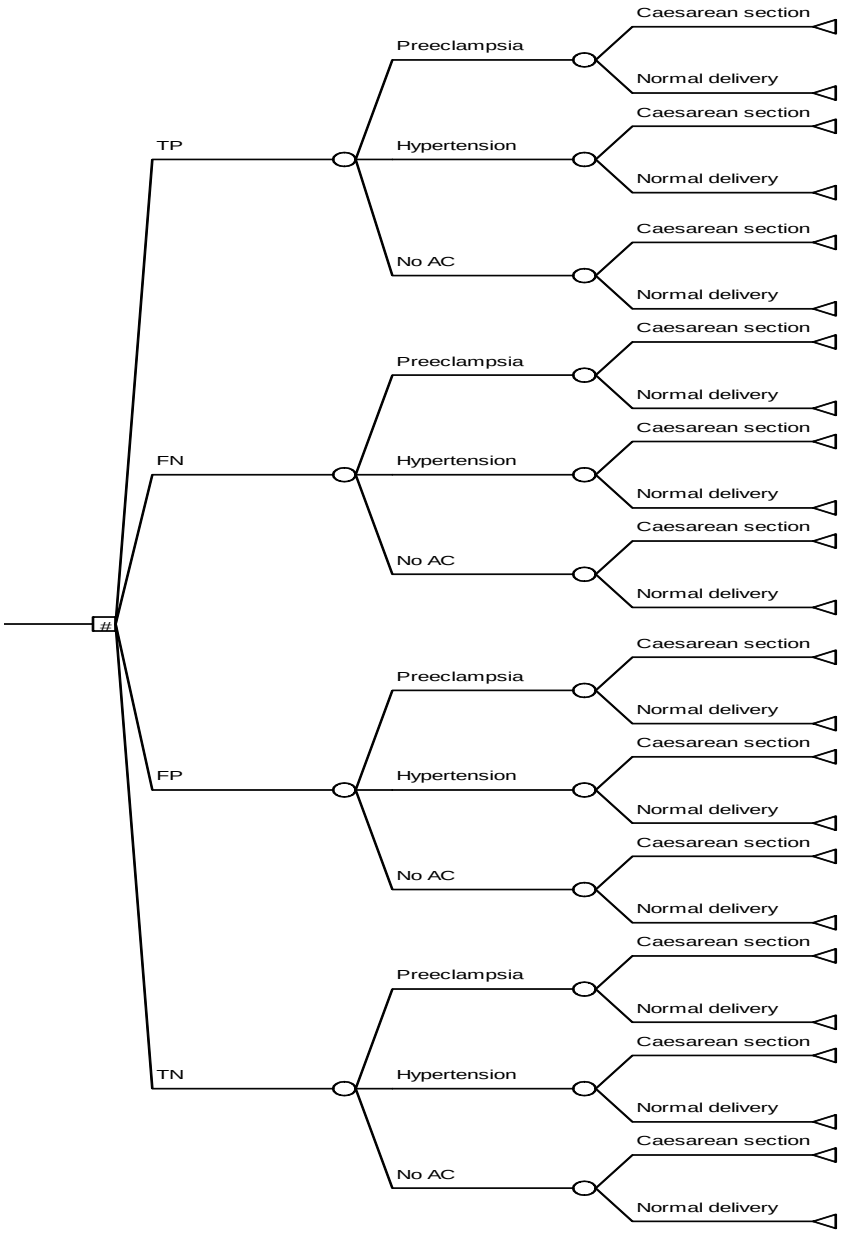


Figure 8.2 Decision tree model for SC in mother

8.3.2 Short term complications in offspring: decision tree structure

Figure 8.3 shows the decision tree of SC in offspring where the chance nodes have four main outcomes including macrosomia, shoulder dystocia, metabolic problems, and no AC. All these SC are taken into account from after delivery until age at one year. This model assumes that SC in offspring are conditionally independent of SC in mothers and therefore no link is drawn between complications in mothers and offspring. Applied to these branches are statistics for the likelihood of the risk of adverse outcomes and the relative risks of adverse outcomes with and without glucose control similar to the SC in mothers.

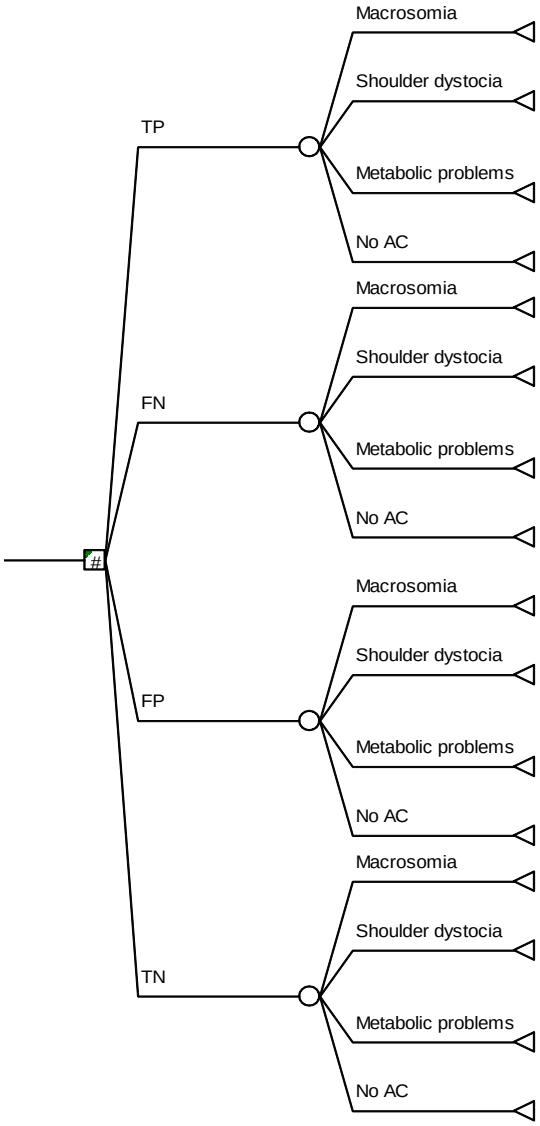


Figure 8.3 Decision tree model for SC in offspring

8.3.3 Short term complications: baseline probabilities

Base-case estimates were obtained from data available in published literature. A search was performed on MEDLINE and EMBASE that included the following terms; gestational diabetes, pregnancy complications, preeclampsia, hypertension, caesarean section, macrosomia, shoulder distocia, metabolic problems, screening test for gestational diabetes, costs, and utility, as addressed in chapter 4. Parameter estimates were determined from systematic reviews, meta-analysis, randomised controlled trials, and perspective cohort studies where available. Retrospective cohort studies, expert opinions or internal data from our institution were used when no other sources of information were available. In addition, a brief presented of short and long term complications has only been made in chapter 4, however this was not clear enough to populate the economic model for GDM. As a result, all baseline probabilities were clearly tabulated in this section including references and probability distributions.

All outcomes from the various studies can be integrated into a single number, and it is assumed that the parameters are drawn from some common prior distribution whose parameters are unknown. If there are multiple trials analysing the same populations then we can pool information from the trials. α is used to denote the number of mothers who have events per study, and n the number of patients per study, while θ denotes the estimated results obtained from pooling the data together. For example, Caesarean delivery is reported in two RCTs (Crowther et al., 2005) (Landon et al., 2009) from control groups. The sum of the values for α divided by the sum of the n values can be pooled to estimate θ . The baseline parameter of caesarean delivery is therefore estimated to be 32.95 (318/965), as shown in Table 8.7. This figure is the baseline risk from control group.

Table 8.7 Pooling the parameter numbers

Variable	Reference	α	β	n	θ
Caesarean delivery	(Landon et al., 2009)	154	301	455	33.85
Caesarean delivery	(Crowther et al., 2005)	164	346	510	32.16
		318	647	965	32.95

The RCTs were undertaken to evaluate the clinical effectiveness of treatment for GDM. Baseline risks used in the model were taken from the control groups in the trials, as in Table 8.7 (32.95). Additionally, to estimate the probability of adverse complication rates with and without treatment, odds ratios were used. Systematic review and meta-analysis of the effects of treatment in women with GDM which present odds ratios are used to calculate probability in the treatment arms. For example, the effect of treatment in women with GDM from meta-analysis reported odds ratios at 0.86 (95% CI 0.72-1.02)(Horvath et al., 2010). This study applied odds ratios to baseline probabilities in

order to estimate the corresponding probability in the treatment arms. This is done by first converting baseline probabilities into odds, multiplying by the odds ratio and converting the resulting odds back into a probability, equalling 0.297 as in the example, and representing probability in the treatment arm. The severity of glucose intolerance in women with GDM is not taken into account for complications. All essential and relevant evidence for the GDM model are addressed in this section. A list of the baseline cases and references are shown in Table 8.8. For mothers without GDM or in healthy pregnancy, risks were obtained from articles that studied a similar type of population as used in this study and converted into probabilities for this group. However, there was a lack of studies that presented the risk of SC for healthy pregnancy with treatment for GDM that involved low side-effects as mentioned in chapter 4. This model therefore assumes that healthy pregnancy carries the same risk of developing SC as healthy pregnancy with treatment for GDM.

Previously, different outcome measures have been used in the economic evaluations of screening tests for GDM in both mothers and offspring, as mentioned in section 6.5.3. In presenting the outcomes for both mothers and offspring, most of the economic evaluation models combined the outcomes (for mothers and offspring) and presented them as one single outcome. Such combined outcomes were observed in the cost-effectiveness studies of screening tests for GDM itself and also in the cost-effectiveness analyses of other interventions (Round et al., 2011) (Werner et al., 2012). Consequently, these QALYs were summed as in other maternal and fetal cost-effectiveness studies (Werner et al., 2011). This section illustrates how to calculate the outcome parameters for SC in both mothers. The outcomes are summed and the results outlined in the following chapter.

Preeclampsia

In mothers, preeclampsia is a common complication that occurs during pregnancy (Hosseing-Nezhad et al., 2007) (HAPO, 2009).This model uses Hyperglycemia and Adverse Pregnancy Outcome (HAPO) data which considers the risk of adverse outcomes associated with various degrees of maternal glucose intolerance in 9 countries. The prevalence of preeclampsia in a healthy pregnancy is 4.80% with a range of 1.4% – 11.4%. This model applies this prevalence regardless of whether the mothers that receive treatment in healthy pregnancy result from TN or FP tests. In mothers with GDM, this model also estimates the probability of preeclampsia for individuals with and without treatment from two RCT's (Crowther et al., 2005) (Landon et al., 2009). The baseline case for preeclampsia in GDM without treatment is calculated by pooling the data from those studies together. 12.34 represents the baseline probability in the non

treatment arm and ranges from 5.5 – 18.0. The odds ratio from the systematic review and meta-analysis of the effect of treatment in GDM was 0.27 and this is used to calculate the risk of preeclampsia for the treatment arm from the baseline risk (Horvath et al., 2010). This gives a risk of 3.63 in the treatment arm with a range between 2.5 - 12.0.

Hypertension

The prevalence of hypertension estimates for normal pregnancy were reported as 5.9 % ranging from 0.7% to 17.7% (HAPO, 2009).This model applies this prevalence regardless of whether the mothers that receive treatment in healthy pregnancy result from TN or FP tests. Pregnancies diagnosed with GDM are at a higher risk of hypertension. Two RCTs were used to estimate the probability of hypertension for individuals with and without treatment (Crowther et al., 2005) (Landon et al., 2009). 13.6% (13.5% -13.7%) represents the baseline in the non treatment arm by pooling the data from those studies together. With a lack of any odds ratio for the effect of treatment in GDM for hypertension, this study assumes the odds ratio of preeclampsia to represent the risk of hypertension for the treatment arm from the baseline risk. The probability of hypertension in the treatment arm is estimated as 4.08% (4.0% - 8.60%).

Caesarean sections

Healthy pregnancy's carry a risk of caesarean section (Xiong et al., 2001) (HAPO, 2009) (Hosseing-Nezhad et al., 2007). In the HAPO study 16%, with a range of between 8.6% and 23.5%, is assumed to be the probability of healthy pregnancy with caesarean section, regardless of whether treatment is received or not, resulting from TN and FP tests. Macrosomia is an indirect adverse outcome, which appears to lead to an increase in caesarean deliveries and instrumental deliveries (forceps and ventouse deliveries). GDM patients who are untreated are more prone to giving birth to infants with a weight of over 4000g and required to have caesarean delivery. Different studies on effective treatment of GDM present similar baseline risk factors in control groups (Crowther et al., 2005) (Landon et al., 2009). The findings of these papers are summarised to indicate the probability in the non treatment arm; approximately 32.9% and ranging from 32.0% to 58.13%. 0.86 is the odds ratio from a systematic review and meta-analysis of the effect of treatment in GDM and this is used to calculate the baseline in this arm (Horvath et al., 2010). By applying the odds ratio to the baseline risk of caesarean sections, 29.7% (26.9% - 31.0%) is calculated to be the probability for the treatment arm.

Macrosomia

In offspring, the risk of macrosomia in healthy pregnancy in three different papers show similar rates (Xiong et al., 2001) (Brody et al., 2003) (Chauhan et al., 2005). Chauhan's review of the prevalence of macrosomia in international papers reported risk in the United Kingdom averaging 9%, ranging from 2% to 10%. This is selected to represent the risk of macrosomia in women without GDM (Chauhan et al., 2005). Macrosomia is one of the most common complications related to GDM. Untreated GDM is more likely to lead to infants being born over weight. 14% - 21% of offspring with macrosomia are born form a mother with GDM (Crowther et al., 2005) (Landon et al., 2009). The pooling of baseline figures from RCTs papers gives 17.89 % with a range of 9.09% - 25.5% and were used to indicate the non treatment arm. In the same meta-analysis study it was reported that the odds ratio of the effect of treatment in GDM with macrosomia was 0.38 (Horvath et al., 2010). Baseline risk in the treatment arm which was calculated based on applying the odds ratio to the baseline risk presented as 7.65% with range from 5.9% - 10.0 5%.

Shoulder dystocia

The prevalence of shoulder dystocia in normal pregnancy is 2.10% (0.3 % - 6.4%), and was used to indicate the baseline for mothers without GDM (HAPO, 2009). A high prevalence of macrosomia can cause birth trauma in the offspring of a mother with GDM, and this can make delivery difficult because one of the offspring's shoulders can get stuck behind the pelvic bone. The studies in effectiveness of treatment in GDM were used to estimate the baseline case in the non treatment arm. The pooling of the baseline figures gives a result of 3.47%, ranging between 3.0% from 4.0% (Crowther et al., 2005) (Landon et al., 2009). 0.4 is the odds ratio from meta-analysis of the benefits and harm of treatment for women with GDM (Horvath et al., 2010). In the treatment arm, the risk of shoulder dystocia is estimated as 1.42% (1.0% - 1.5%), based on applying the odds ratio to the baseline risk.

Metabolic problems

There is a low prevalence of infants with metabolic problems in healthy pregnancy (HAPO, 2009) (Hosseing-Nezhad et al., 2007). The prevalence of metabolic problems seen in the HAPO study was 2.10% (0.3% - 6.4%), and has been selected to indicate the baseline for metabolic problems. The children of mothers with GDM may present with metabolic problems. 5.92% with a range of 5.6% - 20.1%, is used to represent the effectiveness of treatment in control groups after pooling the number of events and total level at risk from the two trial studies presented (Crowther et al., 2005) (Landon et al., 2009). The odds ratio from the systematic review and meta-analysis is reported to be

about 0.4. After applying the odds ratio to the baseline risk, the estimate comes out as 2.46% ranging between 2.0% - 5.30% in the treatment arm.

Table 8.8 The list of the resource's of adverse complications.

Variables	Base case	Range	Details	Data source
Mother complication: Preeclampsia				
No GDM	4.80	1.4-11.4	HAPO study in 9 countries	(HAPO, 2009)
	2.01	1.4-2.8	Cross-sectional study	(Hunking et al., 2001)
GDM without treated	5.50	NA	RTC	(Landon et al., 2009)
	18.00	NA	RTC	(Crowther et al., 2005)
	4.38	1.4-9.9	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
GDM with treated	2.50	NA	RTC	(Landon et al., 2009)
	12.00	NA	RTC	(Crowther et al., 2005)
	4.50	NA	Cross-sectional study	(Bryson et al., 2003)
Mother complication: Hypertension				
No GDM	5.90	0.7-17.7	HAPO study in 9 countries	(HAPO, 2009)
	11.40	NA	Retrospective cohort study	(Xiong et al., 2001)
GDM without treated	13.60	NA	RTC	(Landon et al., 2009)
GDM with treated	8.60	NA	RTC	(Landon et al., 2009)
	4.40	NA	Case control study	(Bryson et al., 2003)
Mother complication: Cesarean section				
No GDM	16.00	8.6-23.5	HAPO study in 9 countries	(HAPO, 2009)
	28.10	26.71-29.52	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
	16.20		Retrospective cohort study	(Xiong et al., 2001)
GDM without treated	33.80	NA	RTC	(Landon et al., 2009)
	32.00	NA	RTC	(Crowther et al., 2005)
	47.13	32.00-58.13	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
GDM with treated	26.90	NA	RTC	(Landon et al., 2009)
	31.00	NA	RTC	(Crowther et al., 2005)
Offspring complication: Macrosomia				
No GDM	9.00	2.0-10.0	Review prevalence	(Chauhan et al., 2005)
	10.00	NA	Review prevalence	(Brody et al., 2003)
	9.00	NA	Retrospective cohort study	(Xiong et al., 2001)
GDM without treated	14.30	NA	RTC	(Landon et al., 2009)
	21.00	NA	RTC	(Crowther et al., 2005)
	15.80	9.09-25.5	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
GDM with treated	5.90	NA	RTC	(Landon et al., 2009)
	10.00	NA	RTC	(Crowther et al., 2005)
Offspring complication: Shoulder dystocia				
No GDM	2.10	0.3-6.4	HAPO study in 9 countries	(HAPO, 2009)
GDM without treated	4.00	NA	RTC	(Landon et al., 2009)
	3.00	NA	RTC	(Crowther et al., 2005)
GDM with treated	1.50	NA	RTC	(Landon et al., 2009)
	1.43	NA	RTC	(Crowther et al., 2005)
Offspring complication: Metabolic problems				
If no GDM	2.10	0.3-6.4	HAPO study in 9 countries	(HAPO, 2009)
	3.90	3.32-4.55	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
GDM without treated	6.80	NA	RTC	(Landon et al., 2009)
	11.49	5.6-20.1	Cross-sectional study	(Hosseing-Nezhad et al., 2007)
GDM with treated	5.30	NA	RTC	(Landon et al., 2009)

The baseline clinical parameters used in the decision tree model are summarised in Table 8.9. The data are binomial, in the form of a baseline probability, and therefore it was decided that Beta distribution was appropriate to represent uncertainty in the probabilistic analysis for TN, FP and FN. On the other hand, in the TP arm, this model uses the odds ratio from the literature review to incorporate the effect of treatment into the model. The appropriate distribution for this is lognormal (Briggs et al., 2006).

Table 8.9 Summary short term complication parameters in the model

Variables	Parameter estimated	Probability distribution	Parameters of distribution		Data source
			Alpha	Beta	
Mother complication: Preeclampsia					
No GDM without treated	0.046	Beta	1,116	24,389	(HAPO, 2009)
No GDM with treated	0.046	Beta	1,116	24,389	(HAPO, 2009)
GDM without treated	0.122	Beta	118	847	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.700	Log normal	SE = 0.159		(Horvath et al., 2010)
Mother complication: Hypertension					
No GDM without treated	0.057	Beta	1,370	24,135	(HAPO, 2009)
No GDM with treated	0.057	Beta	1,370	24,135	(HAPO, 2009)
GDM without treated	0.136	Beta	62	393	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.700	Log normal	SE = 0.159		(Horvath et al., 2010)
Mother complication: Cesarean section					
No GDM without treated	0.171	Beta	3,731	21,774	(HAPO, 2009)
No GDM with treated	0.171	Beta	3,731	21,774	(HAPO, 2009)
GDM without treated	0.333	Beta	318	674	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.860	Log normal	SE = 0.089		(Horvath et al., 2010)
Offspring complication: Macrosomia					
No GDM without treated	0.097	Beta	763	7,854	(Chauhan et al., 2005)
No GDM with treated	0.097	Beta	763	7,854	(Chauhan et al., 2005)
GDM without treated	0.179	Beta	175	803	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.380	Log normal	SE = 0.125		(Horvath et al., 2010)
Offspring complication: Shoulder dystocia					
No GDM without treated	0.014	Beta	311	23,005	(HAPO, 2009)
No GDM with treated	0.014	Beta	311	23,005	(HAPO, 2009)
GDM without treated	0.035	Beta	34	945	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.400	Log normal	SE = 0.325		(Horvath et al., 2010)
Offspring complication: Metabolic problems					
No GDM without treated	0.042	Beta	950	22,366	(HAPO, 2009)
No GDM with treated	0.042	Beta	950	22,366	(HAPO, 2009)
GDM without treated	0.777	Beta	75	904	(Crowther et al., 2005) (Landon et al., 2009)
GDM with treated	0.760	Log normal	SE = 0.241		(Horvath et al., 2010)

8.3.4 Short term complications: cost estimates

All important costs calculated in foreign currencies were converted into pound sterling (£) for use in calculations in the model as mentioned in section 7.5.2. The model also includes the costs of additional outcomes from systematic reviews, meta-analysis and randomised controlled trials on the effect of treatment of GDM on pregnancy. Care costs for maternal and infant complications have been sourced from published literature and cost reports (Barton et al., 2006), as shown in table 8.10. A report of hospital costs, namely the Admitted Patient Care Mandatory Tariff produced by the Health Resource Group (HRG) has been used in this study (NHS TARIFF 06/07, 2007) (NHS TARIFF 10/11, 2011). The cost of a normal delivery is £971.00 (lower percentile at 721 and upper percentile at 1174). In this study, delivery costs without complications are calculated and included in the costs for the care of adverse complications. Regarding caesarean section costs within the report, only caesarean section without complications is considered. Hypertension costs during gestation were not reported. Thus,

hypertension costs are assumed to be the same as preeclampsia costs. All costs that have been used in the decision tree model are summarised in Table 8.10.

Table 8.10 The list of the resource's cost of adverse complications.

Variable	Base case (£)	Range	Details	Reference
Healthy mother				
Normal delivery without complications	971.00	NA	Price yr 2011*	(NHS TARIFF 10/11, 2011)
Mother adverse compilations				
Preeclampsia	261.87	NA	Price yr 2008*	(NHS TARIFF 06/07, 2007)
	1,335.77	1,012-13,184.14	Price yr 2006*	(Barton et al., 2006)
Hypertension	261.87	NA	Price yr 2008*	Assumed as the same price as preeclampsia
Caesarean without complications	1,544.00	NA	Price yr 2011*	(NHS TARIFF 10/11, 2011)
	5,274.45	NA	Price yr 2002**	(Di Cianni et al., 2002)
Offspring adverse compilations				
Macrosomia	623.71		Price yr 2008*	(NHS TARIFF 06/07, 2007)
Shoulder dystocia	683.48		Price yr 2008*	(NHS TARIFF 06/07, 2007)
Metabolic problems	896.45		Price yr 2008*	(NHS TARIFF 06/07, 2007)
	1,318.41	533.70-3,994.87	Price yr 2003*	(Gilbert et al., 2003)

** Hospital cost, ** Reimbursements for hospitalization are base on DRG

The gamma distribution is frequently used to model costs (Briggs et al., 2006). The reason some papers prefer it over the normal is that the normal assumes a symmetric distribution while the gamma allows for a right skewed distribution, which is usually apparent in raw cost data. However, although costs on the raw scale are likely to be right skewed our interest for modelling lies in the distribution of mean costs, and if the sample from which costs are estimated is large enough the distribution of mean costs may be approximately normally distributed even if the underlying distribution is not (Briggs et al., 2006). This is why costs are often modelled using a normal distribution.

As mentioned above, normal delivery costs were found to be £971.00 (lower percentile (15) at 721 and upper percentile (75) at 1174). NHS reference costs however only provide 15th & 75th percentiles of the cost distribution. To fit cost data with the normal distribution, an estimate of the standard error (SE) is needed. Normally the SE can be calculated using the confidence interval, namely (upper-lower) / (2*1.96). The 95% CI covers 95% of the normal distribution and equates to 2 standard deviations (SD) from the mean value. The 15th and 75th percentiles can be used to calculate a 95% CI, and as the difference between the lower (15) and upper (75) percentiles equals 3.2 SD from the mean value, the value for 1SD can then be calculated.

Recalling the cost of a normal delivery, 1SD is calculated as (1174 – 721)/3.2 = 141.60. The cost has a mean of 971.00. The 95% confident interval is therefore

computed by subtracting 2 (SD) from the mean to find the lower value and adding 2 (SD) to give the upper value of the 95% CI, which equal 687.88 and 1254.13, respectively. In cases where no SE is available from data sources, an alternative approach is to use data that present a 95% CI and mean value to make the necessary calculations. Unfortunately, if the cost values do not report both SE and 95% CI, it is common to assume some measure of uncertainty. In this respect, one of the most common approaches in the literature is to assume what is called a coefficient of variation. So for example, a coefficient of variation of 20% would mean that the SE is 20% of the mean. The choice of 20%, 30% or 100% is completely arbitrary. This model assumes the SE is the same value as the mean (Briggs et al., 2006). The unit cost, SE and distribution used in the probabilistic analysis are report in Table 8.11.

Table 8.11 Cost parameters in the model

Cost	Parameter estimate	Probability distribution	SE	Data source
Healthy mother				
Normal delivery without complications	971.00	Normal	144.45	(NHS TARIFF 10/11, 2011)
Mother adverse compilations				
Preeclampsia	261.87	Normal	127.55	(NHS TARIFF 06/07, 2007)
Hypertension	261.87	Normal	261.87	(NHS TARIFF 06/07, 2007)
Caesarean	1544.00	Normal	346.94	(NHS TARIFF 06/07, 2007)
Offspring adverse compilations				
Macrosomia	623.71	Normal	623.71	(NHS TARIFF 06/07, 2007)
Shoulder dystocia	683.48	Normal	683.48	(NHS TARIFF 06/07, 2007)
Metabolic problems	896.45	Normal	896.45	(NHS TARIFF 06/07, 2007)

8.3.5 Short term complications: QALY over one year in mothers

The term “quality of life” and more specifically health related quality of life (HRQoL) refer to the physical, psychological, and social domains of health, which cover all WHO aspects, as mentioned in section 2.5.1 (Marcia and Donald, 1996). To estimate the quality adjusted life year, it is necessary to quality-adjust the period for which the average patient is alive within the model using an appropriate utility or preference score. As mentioned in the above section, GDM have various SC events in both mother and offspring. The GDM SC in mothers required for this economic model includes healthy pregnancy, preeclampsia, hypertension, and caesarean section.

In order to determine utility for individual GDM events, estimates of utility were obtained from various data in the published literature. Currently, there are many instruments available to measure HRQoL. The most common technique for measuring the HRQoL in pregnant women is 36-Item Short Form (SF-36), which is a widely used generic health related quality of life instrument which assesses eight aspects of health

status including physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health (Symon, 2003) (Brazier et al., 2002). However, SF-36 cannot be used in cost effectiveness analysis using cost per QALY because it is not preference based. Instead, preference-based measures such as the EQ-5D are generally used (Rowen et al., 2009) (NICE, 2013). In addition, EQ-5D is a standardised instrument which is applicable to a wide range of health conditions and treatments, and used as a measure of health outcomes. This instrument provides a simple descriptive profile and a single index value for health status that can be used in the economic evaluation of health care. Moreover, EQ-5D had specially been designed to complement other quality of life measures such as the SF-36, NHP and disease-specific measures (SIGN., 2002). Mapping from health status measures onto generic preference-base measures, which was studied by Rowen and colleagues, is being utilised when health state utility values are not available for economic evaluation (Rowen et al., 2009). SF-36 dimension scores are mapped on to the EQ-5D index using a published algorithm (Rowen et al., 2009). In this model, both mother and offspring, calculated their utility over one year (52 weeks). The calculation of the utility over one year period is discussed in this section.

Before introducing the utility in the mother, it is important to consider the duration of pregnancy, as shown in Figure 8.4. The duration of pregnancy in humans is about 9 months (266 days / 38 weeks) from the time of fertilization until birth. In obstetrics, it is considered to begin on the first day of women's last normal menstrual period prior to fertilization, thus beginning approximately 280 days (40 weeks) before birth. Those 40 weeks, counting from the first day of a woman's last normal period, are grouped into three trimesters. Normally, childbirth occurs about 38 – 40 weeks after conception or the last normal menstrual period. The first trimester starts from week 1 to week 12, the change in maternal hormones affects almost every organ system in the mothers body. Consequently, week 13 to week 28 is the second trimester for gestation, in which the maternal abdomen will expand as the baby continues to grow and in which the mother feels their baby beginning to move. Lastly, the third trimester starts from week 29 to week 40, all of the discomfort from the second trimester continue in addition to many more symptoms. This is due to the baby getting bigger and consequently putting more pressure on the mother's organs. Furthermore, pregnancy is considered "at term" when gestation has lasted 37 complete weeks, but is less than 42 weeks gestational age. Likewise, "full term delivery" refers to gestation having lasted 40 weeks from the first day of the mother's last menstrual period. Events before completion of 37 weeks (259 days) are considered preterm. Delivery after week 42 (294 days) is considered postterm of pregnancy. The expected utility of life for the mother in this model is calculated within 1 year (52 weeks). This model assumed all healthy pregnancies give birth at 40 weeks

"full term". The duration of one year breaks down into two periods of time, which includes 40 weeks during gestation and 12 weeks after delivery. Utility calculation is shown in the following section.

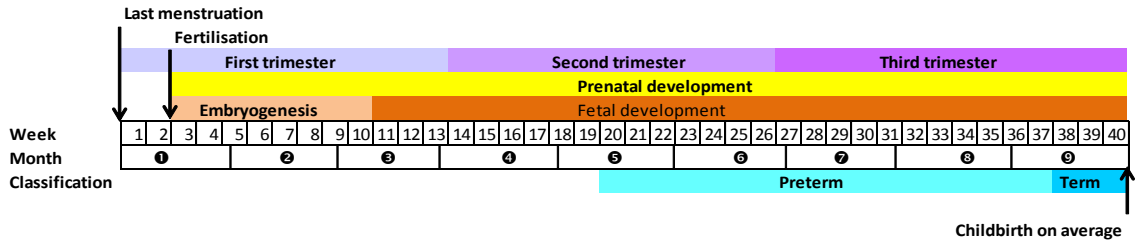


Figure 8.4 Overlay of trimester, gestational week, and term of pregnancy

Utility of healthy mothers

The baseline utility of life refers to healthy mothers with vaginal delivery, calculated over 1 year. During the progress of pregnancy hormonal and organ specific changes impact physical as well as the mental wellbeing of the mother. A study in Switzerland, in 2005, has evaluated the impact of pregnancy on health related quality of life (Forger et al., 2005). HRQoL of healthy pregnant women at 29 years old was measured by SF-36 and carried out at the following times: one at the first trimester (between weeks 9 and 11), one at the second trimester (between 20 and 22), one at the third trimester (between weeks 30 and 34), and once each at 6 and 12 weeks postpartum. All eight aspects of SF-36 in the Swiss study were mapped onto EQ-5D to present the EQ-5D tariff in healthy pregnancy for each trimester and at 6 and 12 weeks postpartum. Then, multiplying the duration of time spent in a "health state" with the EQ-5D tariff, during gestation, within 40 weeks, the utility of healthy mothers was presented for each trimester as 0.97, 0.96, and 0.88, respectively. At 12 weeks postpartum, the utility of life for a vaginal delivery based on the first six weeks is a utility of 0.85. The remainder of the 6 weeks of normal delivery for recovery is a utility of 0.98. The following section will demonstrate how to calculate QALY over one year from the above figure.

There were two steps to calculate QALY over one year of health pregnancy. Firstly, the utility over the gestational period of healthy pregnancy is calculated simply by multiplying the duration of time spent in a health state (40 weeks) by the HRQoL weight (utility score), as shown in Table 8.12. For example, in the 1st trimester, utility was 0.97. The utility of the gestational period is calculated by multiplying the utility for each trimester (0.97) by the proportion of weeks within the 40 week period, namely 13.3. The 1st trimester utility was calculated to be 0.323 and the mean utility for the gestational period was 0.934, calculated by summing the utilities for each of the three trimesters. Secondly, in the postpartum period of 12 weeks, the utility for the first 6 weeks was 0.85. As before, the utility for this first period (0.85) was multiplied by its proportion (6 weeks)

within the postpartum period of 12 weeks. This is calculated as 0.85*(6/12), equalling 0.425. The mean utility is then calculated by summing the utilities for the 6 and 12 week periods, resulting in 0.915.

Table 8.12 Utility calculation over gestational and postpartum periods

Gestational period (40 weeks)			Postpartum period (12 weeks)		
Duration	Calculation	Utiliy	Duration	Calculation	Utiliy
(1-13 weeks)1st	0.97*(13.3/40)	0.323	6 weeks	0.85*(6/12)	0.425
(13 - 28 weeks) 2nd	0.96*(13.3/40)	0.319	12 weeks	0.98*(6/12)	0.490
(29 - 40 weeks) 3rd	0.88*(13.3/40)	0.293			
Mean utility		0.934	**Mean utility**		0.915

The utility over one year (52 weeks) of healthy pregnancy is calculated in this section. The gestation utility was calculated to be 0.719 by multiplying the utility calculated in the previous section (0.934) by the proportion of the year (52 weeks) that the gestation period covers (40 weeks). Likewise, the utility of the post postpartum period (0.915) was calculated to be 0.211, calculated by multiplying the utility by the proportion of time within the one year period. The mean utility for a healthy mother with vaginal delivery is 0.930 over 1 year that includes both gestation and postpartum periods, as shown in Table 8.13.

Table 8.13 Utility calculation over 1 year

Utility over 1 year		
Duration	Calculation	Utiliy
Gestation (40 weeks)	0.934*(40/52)	0.719
Postpartum (12 weeks)	0.915*(12/52)	0.211
Mean utility		0.930

Figure 8.5 illustrates the utility of life in healthy pregnancy over 1 year. This figure shows the utility (y-axis) for different periods of time (x-axis) covering both gestational and postpartum periods in terms of weeks. The EQ-5D tariff in healthy pregnancy for each trimester and at 6 and 12 weeks postpartum are plotted. The red line is represents the utility over a 1 year period which breaks down into three trimesters during gestation and every 6 weeks during postpartum. The purple line represents the mean utility of healthy pregnancy over 1 year at 0.93.

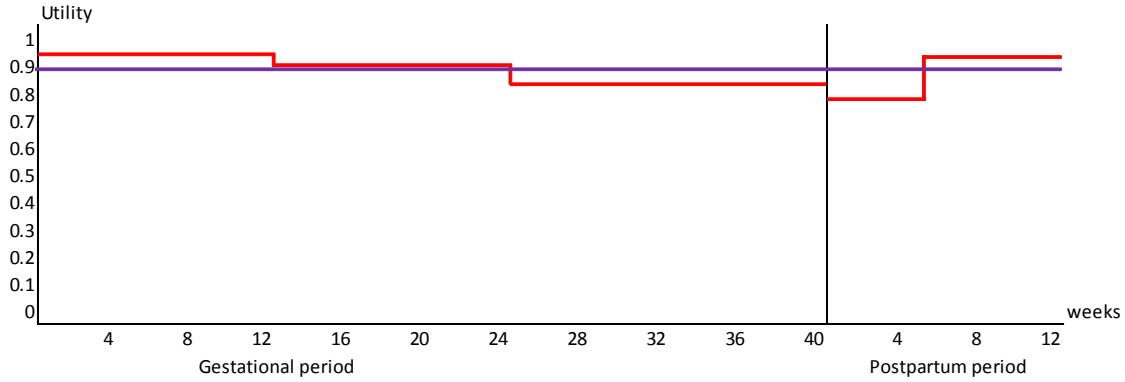


Figure 8.5 Overlay of healthy pregnancy over 1 year

Utility with preeclampsia

Preeclampsia usually develops sometime after the 20th week of gestation and gets better within 6 or 12 months after postpartum (Wagner, 2004). Preeclampsia symptoms include severe headache, problems with vision, vomiting and sudden swelling of the face, hands or feet and severe pain just below the ribs. Treatment with oral medication as first-line can improve maternal quality of life (NICE, 2011). The severity of preeclampsia can cause complications for both mother and offspring. There are few studies investigating quality of life in mothers with preeclampsia. However, of the studies that have been done, most use SF-36. In the Netherlands, Hoedjes studied the HRQoL in pregnancy for ages over 18 years who had complications by preeclampsia. The HRQoL of Hoedjes's study stated that women who experienced severe preeclampsia had a lower postpartum HRQoL than those pregnancies that had mild preeclampsia. All participants completed the SF-36 at 6 weeks and 12 weeks postpartum after mild and severe preeclampsia (Hoedjes, 2011). This model, then maps the SF-36 in Hoedjes study to the EQ-5D as described above. In particular for women that have experienced severe preeclampsia the EQ-5D tariff at 6 and 12 weeks postpartum were estimated to be 0.765 and 0.874, respectively. In this model, all women diagnosed with preeclampsia in the 20th week of gestation are assumed to have severe preeclampsia without failure in treatment. The benefits of the treatment of preeclampsia were also ignored.

As no utility for the gestation period was available, to calculate the utility of severe preeclampsia, the utility of life is separated into three periods; two 20 weeks periods followed by the 12 week postpartum period, as shown in figure 7.6, and a calculation is performed for the postpartum period using the EQ-5D tariff from Hoedjes study. The utility of life for postpartum with severe preeclampsia was based on the first 6 weeks with a utility of 0.765 and the remaining 6 weeks of recovery at a utility of 0.874, as illustrated by the red line in Figure 8.6. Therefore, the mean utility of postpartum following severe preeclampsia was 0.82, as presented by the broken blue line in the postpartum period.

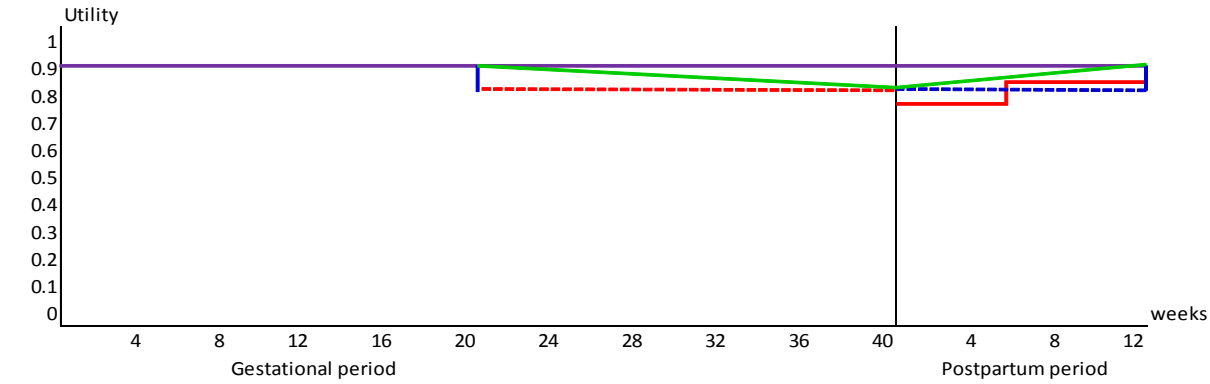


Figure 8.6 Overlay of utility pregnancy with preeclampsia over 1 year

The mean utility between 0 and 20 weeks gestational was then calculated, with the assumption that all women remain in healthy pregnancy before being diagnosed with preeclampsia and have a utility of 0.93, as represented by the purple line in the figure. This gives a result of 0.465 (0.93*0.5) for the mean utility over the first 20 week period.

The utility of the second 20 week period of gestation with preeclampsia was then calculated. A purple line was drawn on the figure to represent healthy pregnancy over the 52 week period, and at week 20 utility was considered to be 0.93 falling to a utility of 0.82 at week 40, as shown by the green line in the figure. The average utility during this gap period is given by adding the highest and lowest utilities and dividing by 2, i.e. (0.93 + 0.82)/2 = 0.874. The utility for this section is therefore 0.874 multiplied by 0.5 which equals 0.437. Therefore, the mean utility over the 40 week period for women with severe preeclampsia was 0.902. Utility at 40 weeks (0.902) and at 12 weeks (0.82) are multiplied by the proportions of time for both the gestational and postpartum periods in the year. This gives a mean utility of 0.88 over 1 year for women with severe preeclampsia.

Utility with hypertension

Hypertension is a common complication in pregnancy, and is defined as mothers diagnosed with a blood pressure of 140/90 or above. On the other hand, preeclampsia is a syndrome which includes hypertension and loss of protein in the urine (Xiong et al., 2002). There is a slight difference in blood pressure between hypertension in pregnancy and preeclampsia (Hermida et al., 2000). However, as no data for the utility of life in pregnant women who have hypertension was available, this model uses the HRQoL for mild preeclampsia in Hoedjes's study to represent the utility of life for hypertension (Hoedjes, 2011).

As in the previous section SF-36 was mapped into EQ-5D and the HRQol of women with mild preeclampsia at 6 weeks and 12 weeks postpartum were estimated to be 0.893 and

0.941, respectively. The mean utility at 12 weeks of hypertension was 0.88. To calculate the utility of hypertension over 1 year it is best to separate the utility of life into three periods as for preeclampsia. Figure 8.7 illustrates those three periods. The same technique that was used to calculate utility in preeclampsia was again used to estimate utility in hypertension over the gestation (40 weeks) and postpartum (12 weeks) periods. During the first 20 weeks, all pregnancy is assumed to be healthy. 0.93 is used to represent the health of mothers during this period. After 20 weeks gestation (utility at 0.93) the utility starts to fall until 40 weeks (utility 0.88) as shown by the green line. The average utility is then given by the sum of the highest value and the lowest value divided by 2; (0.93+0.88)/2, calculated as 0.905. The utility of the first 20 weeks (0.93) and the following 20 weeks (0.905) are multiplied by the proportion of time within the 40 week period, equalling 0.464 and 0.453, respectively. Therefore, the mean utility of hypertension in pregnancy at 40 weeks was 0.917. The mean utility over 1 year for women with hypertension was 0.910, calculated by multiplying the utility for both the 40 weeks of gestation (0.917) and the 12 weeks postpartum (0.88) with their relative proportions of the 52 week period.

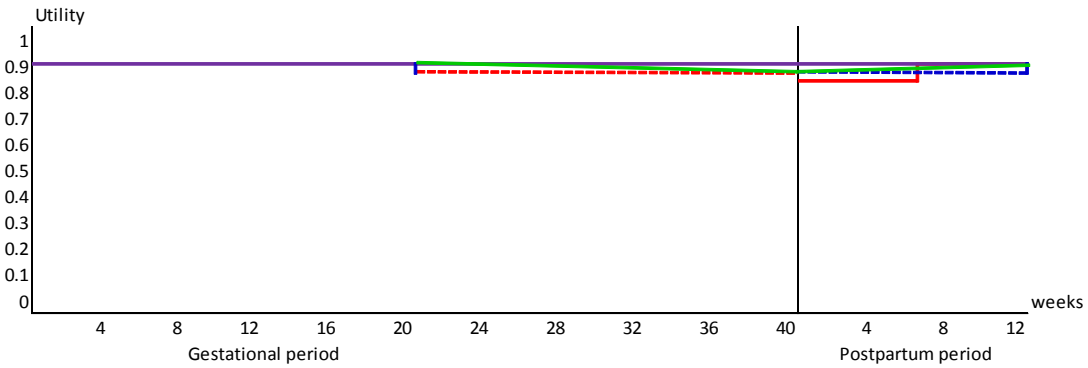


Figure 8.7 Overlay of the utility of pregnancy with hypertension over 1 year

A comparison of the utility of hypertension in general with/ without hypertension and pregnancy with/ without hypertension is discussed in this section. The systematic review of HRQoL in hypertension stated that the quality of life of individuals with hypertension is slightly worse than that seen in healthy people. The meta-analysis showed lower quality of life in both the MCS score (-1.68: 95% confidence interval -2.14 to -1.23) and the PCS score (-2.43: 95% confidence interval -4.77 to 0.88) (Trevisol et al., 2011). This study calculates the figure for HRQoL from a population based study in Sweden by using the SF-36 from the study to represent the baseline utility in the general population and also compares the HRQoL of people with and without hypertension by age and sex. The results obtained for HRQoL in women are used to present the quality of women in the general population with and without hypertension. After mapping the SF-36 into EQ-5D, the HRQol of women with and without hypertension were 0.73 and 0.86, respectively. The difference in HRQoL in both groups is 0.13 per utility. Similarly, the utility difference

in hypertension in pregnancy with and without hypertension in this calculation was 0.2 (0.93 – 0.91).

Utility with caesarean delivery

Caesarean delivery might increase the risk of surgical intervention and problems resulting from hospitalisation. Consequently, this affects the quality of life for women following delivery. This model assumes all pregnant mothers experience caesarean delivery without complications at week 40 of gestation. There are only a few studies that review quality of life following caesarean section, all of them present the HRQoL using SF-36. Postnatal quality of life in women after caesarean section was studied by Torkan and his colleagues in Iran (Torkan et al., 2009). Quality of life of the women in the study was measured using SF-36 for two periods of time (time 1: 6 to 8 weeks after delivery; time 2: 12 to 14 weeks after delivery). Postpartum quality of life was improved from period 1 to period 2. After mapping, the EQ-5D tariff for women with experience of caesarean section at 6 and 12 weeks was 0.77 and 0.82, respectively (Torkan et al., 2009). The postpartum quality of life was calculated from Token's study. The expected quality of life for a caesarean delivery without complications was based on the first 6 weeks with a utility of 0.77, followed by a utility of 0.82 for the following 6 week, as shown in Figure 8.8. Therefore, the mean utility in the postpartum period (12 weeks) of caesarean section was 0.80.

To calculate the utility of caesarean sections at 40 weeks of gestation, this model assumed that all pregnant women during the gestational period are healthy. The utility of healthy pregnancy was calculated as 0.93. Therefore, the utility over 1 year was calculated by multiplying the utility estimated for each period by the proportion of each period in 52 weeks. The utility for caesarean delivery over one year was found to be 0.903.

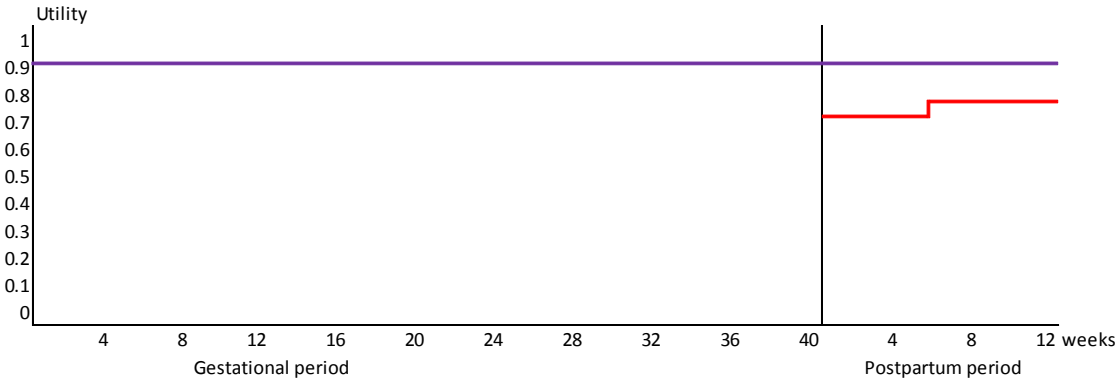


Figure 8.8 Overlay the utility of pregnancy with caesarean section over 1 year

8.3.6 Short term complications: QALY over one year in offspring

The GDM SC for offspring required for this economic model included macrosomia, shoulder distocia and metabolic problems. Before calculating the utility of AC in offspring, it is necessary to introduce the short and long term complications after having macrosomia, shoulder dystocia, metabolic problems and NICU. The beginning of this section discusses these short and long term complications in offspring and the details are summarised in Table 8.14.

Macrosomia has long been associated with an increased risk for mother and baby, both during pregnancy and after childbirth, as mentioned in section 4.5. Possible maternal complications of foetal macrosomia might include labor problems, genital tract LCerations, bleeding after delivery, and uterine rupture. Possible complications of foetal macrosomia for babies might include higher than normal blood sugar levels, childhood obesity, metabolic syndrome and shoulder dystocia. Macrosomia impacts on the offspring in the long term more than short term in terms of complications. It is assumed that all mothers who give birth to babies weighing over 4500g, receive caesarean sections. This study therefore ignores possible maternal complications after delivery.

Shoulder dystocia occurs when a baby's shoulder is trapped during delivery and requires prompt medical attention by delivery room staff or NICU. 20% of babies suffer some sort of injury either temporary or permanent following shoulder dystocia delivery. Usually, both mother and baby do well and have no permanent complications. In most cases injury to nerves in the shoulder, arms and hand, problems go away in 6 – 12 months. Regarding serious complications for the newborn, shoulder dystocia can cause brain damage, cerebral palsy, permanent brachial plexus injury and bone fracture (namely of the clavicle and/or the humerus), contusions and LCerations. These conditions can be mild, disappearing in a short period of time, or severe conditions that affect a child for the whole of his or her life. They can be significant and debilitation and can have a serious impact on the child's quality of life along with that of his or her parents. Babies that have arm paralysis due to shoulder dystocia are often told that within a matter of months or a year the arm's function will return to normal. Paralysis does not go away in more severe cases, however a full recovery can be expected in cases where nerve injury has been less severe. The injury requires physical therapy and rehabilitation and there is loss of quality of life.

Neonatal metabolic problems result in increased admissions to Neonatal Intensive Care Units (NICU). Jensen et al. showed in their study that about 46.2% of offspring with GDM mothers were admitted to a neonatal unit compared to 11.9% of

mothers without GDM (Jensen et al., 2000). Neonates require intensive care treatment much more than an adult. Approximately 70% of neonates stay 20 days or more in the NICU. This prolonged hospitalization suggests that an infant's intensive care experience could inform strategies to minimize the burdens associated with NICU interventions to improve HRQoL of the child. On the other hand, long term complications for offspring with GDM mother include type 2 diabetes mellitus (DM).

Table 8.14 Short and long term complications in offspring AC events

Offspring AC events	Short term complications	Long term complications
Macrosomia	-Shoulder dystocia -Genital tract LCerations -Bleeding after delivery -Uterine rupture -3 days in NICU	-type 2 DM
Shoulder distocia	-Nerves of the shoulder, arms and hand (the problem go away in 6 – 12 months) - 1 month in NICU	- Brain damage - Cerebral palsy - Permanent brachial plexus injury - Bone fracture -Type 2 DM
Metabolic problems	1 week in NICU	-Type 2 DM

For cases that involve an adverse outcome following birth, this model calculated the utility for offspring over 1 year (52 weeks) after delivery, including the neonatal period (birth to 1 month) and infancy (1 month to 1 year). It is assumed that all offspring are delivered at 40 weeks of gestation. The utility of life for offspring during the 12 weeks after delivery is calculated based on the consideration of short term complications only and for the remaining 40 weeks it is assumed that all offspring return to prefect health. Neonatal quality of life cannot readily be studied. The utility in offspring during this 12 week period has not been determined in previous studies. Therefore the mean utility of macrosomia, shoulder dystocia and metabolic problems in this model has been assumed based on the study by Tan and colleagues of the cost-effectiveness of external cephalic version for term breech presentation and expert opinions (Tan et al., 2010). The study by Tan and his colleagues, assumed the mean utility in NICU's to be 0.20 with a utility of 0.66 for the subsequent 2 weeks. For the rest of the period, the utility was 0.77. This model has adapted the above figures proved by experts in the field as follows.

Utility with Healthy offspring

As mentioned above, there are no studies that review quality of life during the neonatal period. Perfectly healthy offspring are assumed to be children born by vaginal delivery without complications. The mean utility for healthy neonatal offspring and infants over the 40 week period is equal to 1.

Utility with macrosomia

Macrosomia with short term complications are used to estimate utility in offspring in the 12 weeks after birth. The utility of macrosomia was based on 3 days in NICU at utility of 0.2. The following 2 weeks post-delivery utility was 0.66, with a utility of 0.77 for the following 8 weeks until there was a return to perfect health. The mean utility of offspring with macrosomia at 12 weeks was 0.746, calculated from the utilities at the three different stages (0.2,0.66 and 0.77) multiplied by their relative proportion of the 12 week period, 0.5, 2.5 and 9 weeks respectively. The model assumes all offspring with macrosomia subsequently return to perfect health at utility of 1. Therefore, the utility of macrosomia over 1 year was 0.94, calculated by taking the utility for the 12 week period (0.746) multiplied by its proportion of 52 weeks and adding it to the utility at 40 weeks (1) multiplied by its proportion of the 52 week period. Utility over a 1 year period for offspring with macrosomia is displayed in Figure 8.9.

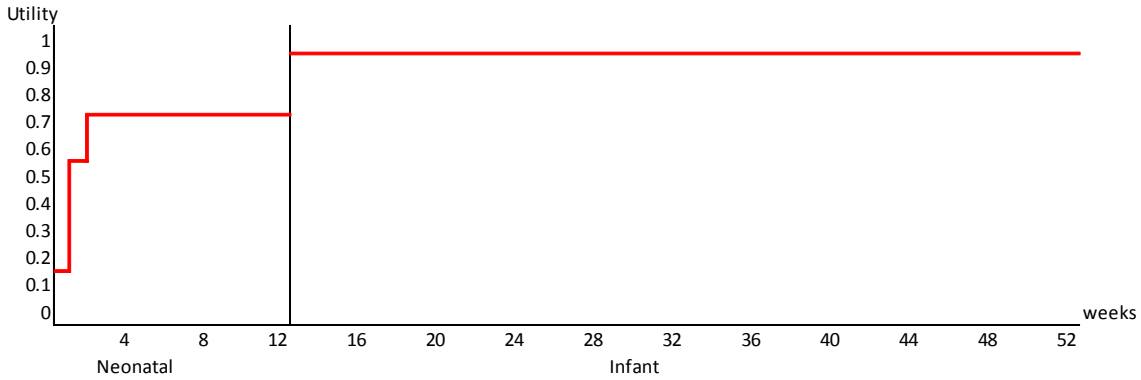


Figure 8.9 Overlay the utility of offspring with macrosomia over 1 year

Utility with shoulder dystocia

Based on the study by Tan et al. and other expert opinion it was assumed that offspring with shoulder dystocia and short term complications require a 4 week stay in NICU with a utility of 0.20 and that the following 8 weeks have a utility of 0.77 (Tan et al., 2010), as shown in figure 8.10. The mean utility over 12 weeks for shoulder dystocia was 0.58. After 12 weeks delivery, all offspring return to perfect health at utility 1. Therefore, utility over 1 year with shoulder dystocia was 0.90, calculated from the utility at 12 weeks and at 40 weeks multiplied by their proportion of the 52 week period, 12 and 40 weeks, as shown in Figure 8.10

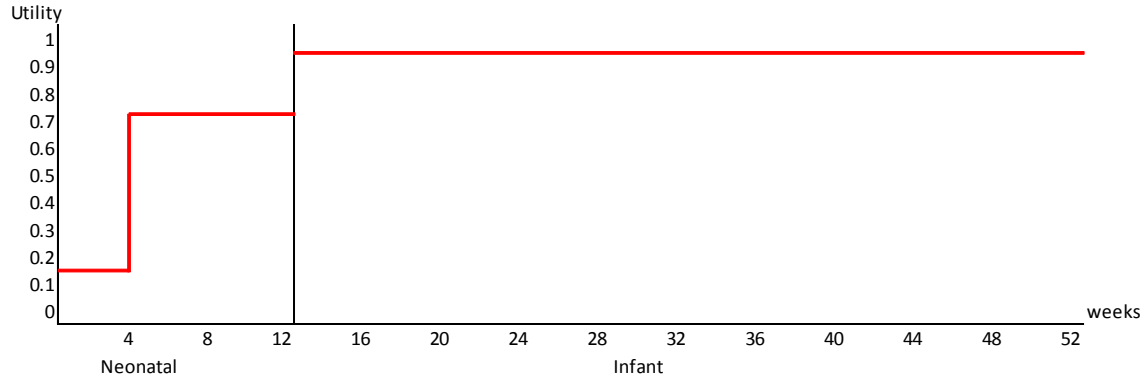


Figure 8.10 Overlay the utility of offspring with shoulder dystocia over 1 year

Utility with metabolic problems

The expected quality of life for offspring with metabolic problems and no complications was based on the first weeks at NICU with a utility of 0.2 and the remaining 11 weeks recovery period with a utility of 0.77. (Tan et al., 2010). The mean utility when there are metabolic problems was 0.72 within 12 weeks post-delivery. After 12 weeks, offspring return to perfect health at a utility 1. The mean utility over 1 year for offspring with metabolic problems and no complications was 0.94 at 52 weeks, as shown in Figure 8.11.

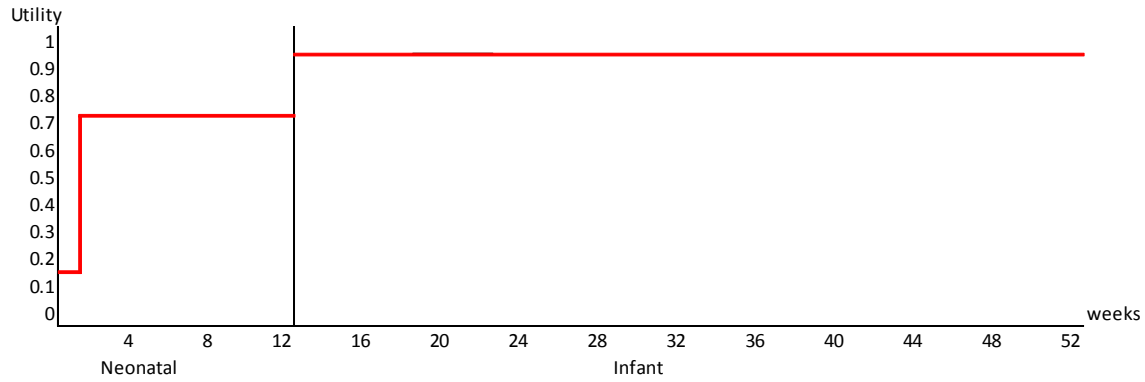


Figure 8.11 Overlay the utility of offspring with metabolic problems over 1 year

Gamma distributions were set on disutility (disutility = 1 – utility) so that the distribution was constrained within the interval zero to infinity, allowing very low and even negative utility values. The mean utility value and corresponding SE were reported form the literature, and therefore, the method of moment for gamma was used to calculate the α and β parameter to fit the Gamma distribution, as shown in Table 8.15.

Table 8.15 Short term adverse complication utility parameters in model

Utility	Point estimate	Probability distribution	Parameters of distribution		Data source
			Alpha	Beta	
Healthy pregnancy					
1st trimester	0.97	Gamma	34.65	1.07	(Forger et al., 2005)
2nd trimester	0.96	Gamma	50.37	2.10	(Forger et al., 2005)
3rd trimester	0.88	Gamma	76.24	10.40	(Forger et al., 2005)
6 weeks postpartum	0.85	Gamma	117.69	20.77	(Forger et al., 2005)
12 weeks postpartum	0.98	Gamma	52.80	1.08	(Forger et al., 2005)
Preeclampsia					
6 weeks postpartum	0.76	Gamma	276.69	85.00	(Hoedjes, 2011)
12 weeks postpartum	0.87	Gamma	225.45	32.50	(Hoedjes, 2011)
Hypertension					
6 weeks postpartum	0.82	Gamma	136.31	29.32	(Hoedjes, 2011)
12 weeks postpartum	0.93	Gamma	38.45	32.096	(Hoedjes, 2011)
Caesarean section					
6 weeks postpartum	0.77	Gamma	153.33	44.01	(Torkan et al., 2009)
12 weeks postpartum	0.82	Gamma	147.02	32.06	(Torkan et al., 2009)
Offspring					
Macrosomia	0.93	Gamma	1568.78	79.58	(Tan et al., 2010)
Shoulder dystocia	0.88	Gamma	157.48	16.90	(Tan et al., 2010)
Metabolic problems	0.92	Gamma	163.22	11.17	(Tan et al., 2010)

8.3.7 Impact of GDM events (events decrements)

The HRQol score is a heath utility score which uses weight life years to generate quality adjusted life years (QALY). Individuals are rarely in perfect health, with perfect health equating to 1 on the scale. A reduction in utility from perfect health is called a utility decrement. Various events and complications have the potential to reduce HRQoL. This model calculated the utility for seven GDM events and assumed that there is a possibility for all pregnant women to experience these events. In the previous section, perfect health has a utility of 0.93 for mothers and 1 for offspring. For these results, the utility of AC events were subtracted from the utility for perfect health in both mothers and offspring. This event decrement was presented over one year. The GDM event decrements rank health from 0.02 (hypertension) to 0.10 (shoulder dystocia), as shown in Table 8.16. The greatest reductions in HRQoL were found with shoulder dystocia in offspring and preeclampsia in mothers.

Table 8.16 GDM event decrements

Variables	QALY over 1 year	Utility drecrement at 1 year
Mother		
Healty pregnancy	0.930	
Preeclampsia	0.883	0.05
Hypertension	0.907	0.02
Caesarean	0.903	0.03
Offspring		
Health offspring	1.000	
Macrosomia	0.941	0.06
Shoulder dystocia	0.903	0.10
Metabolic problems	0.936	0.06

In previous cost-effectiveness studies of GDM screen tests, different outcome measures have been used in the economic evaluation of GDM. The utility decrement for events was not presented in any of these studies. This model provides comprehensive and nationally representative estimates of the impact of a wide range of GDM events on preference weighted HRQoL. These estimates can be used in burden of disease studies to estimate the quality of life for mothers and offspring with GDM. The HRQoL reductions serve as utility decrements that can be used in economic evaluations to quantify the impact of interventions designed to prevent or lessen the quality of life impact of GDM.

8.3.8 Expected costs and QALY for short term complications over 1 year for mother

This model assigns a probability value to each outcome branch. The sum of the probabilities for all outcome branches of a single chance node must equal one, as shown in column probability in Table 8.17. Once, the substituted actual numerical variables in each branch apply then the possible decisions by a process called folding back (multiplying associated probability by the monetary and QALY values at each terminal node) can be used. The sum of those processes is called the expected value. This model uses the folding back process in the SC decision tree to present costs and QALY in TP, FN, FP and TN.

Consequence care costs for GDM are calculated by weighting the costs of SC care (costs column) and the baseline parameters of SC (probability column), as shown in Table 8.17. The SC care costs for mothers who have preeclampsia with caesarean section is the highest cost (£1,750.53). The total cost in each arm of TP, FN, FP, and TN were £1,174.11, £ 1,209.29 and £1,086.55, and £1,086.55, respectively.

Expected outcomes of QALY are based on the summation of the pathway values weighted by the probability in each pathway. Utility decrement was weighted by the probability for each AC event. For example, to calculate expected outcomes for preeclampsia in the treatment group, the baseline case in the previous section was used. In the disease arm utility decrement at 0.90 (0.94 – 0.04) was multiplied by the probability of women with GDM in the treatment group that have experienced preeclampsia. Similar calculations were made or the hypertension arm. On the other arm, the utility of healthy mothers (0.94) were multiplied by 1 minus the probability of women with GDM in the treatment group that have experienced preeclampsia and hypertension.

However, healthy pregnancy with treatment for GDM involves low side-effects as mentioned in chapter 4. Various treatments for type 2 DM may differ in ways that have an impact on patient preference and HRQoL. Louis and colleagues studied the utility or disutility of diabetes medication-related attributes (weight change, gastrointestinal side effect and fear of hypoglycaemia in patient with type 2 DM age between 30 and 75 years old in England and Scotland by using the EQ-5D (Matza et al., 2007). In their study, the mean utility of diabetes without complications was 0.89 and the mean utility for patients with oral medication only was 0.87. The different between those utilities is 0.002. This model therefore assumes this number as the disutility of treatment for GDM in healthy pregnancy. The disutility of 0.002 is subtracted from the utility in the FP arm, as shown in Table 8.17.

Table 8.17 Expected costs and QALY over one year in mother

Short term adverse complications	Probability	Costs(£)	Expected cost(£)	Utility	Expected Utility
True positive (TP)					
Preeclampsia+Cesarean	0.026	1,750.53	46.21	0.884	0.0233
Preeclampsia+Normal delivery	0.062	1,232.87	77.00	0.891	0.0557
Hypertension+Cesarean	0.030	1,750.53	51.72	0.908	0.0268
Hypertension+Normal delivery	0.070	1,232.87	86.18	0.914	0.0639
NoAC+Cesarean	0.241	1,488.66	359.00	0.923	0.2226
NoAC+Normal delivery	0.571	971.00	553.99	0.930	0.5305
	1.000		1174.11		0.9229
False negative (FN)					
Preeclampsia+Cesarean	0.040	1,750.53	70.54	0.884	0.0356
Preeclampsia+Normal delivery	0.082	1,232.87	101.08	0.891	0.0731
Hypertension+Cesarean	0.045	1,750.53	78.60	0.908	0.0408
Hypertension+Normal delivery	0.091	1,232.87	112.64	0.914	0.0835
NoAC+Cesarean	0.244	1,488.66	363.73	0.923	0.2255
NoAC+Normal delivery	0.497	971.00	482.71	0.930	0.4623
	1.000		1209.29		0.9208
False positive (FP)					
Preeclampsia+Cesarean	0.008	1,750.53	13.73	0.884	0.0068
Preeclampsia+Normal delivery	0.038	1,232.87	46.75	0.891	0.0330
Hypertension+Cesarean	0.010	1,750.53	17.03	0.908	0.0086
Hypertension+Normal delivery	0.047	1,232.87	57.99	0.914	0.0421
NoAC+Cesarean	0.154	1,488.66	228.93	0.923	0.1419
NoAC+Normal delivery	0.744	971.00	722.13	0.930	0.6915
	1.000		1086.55		0.9240
True negative (TN)					
Preeclampsia+Cesarean	0.008	1,750.53	13.73	0.884	0.0069
Preeclampsia+Normal delivery	0.038	1,232.87	46.75	0.891	0.0338
Hypertension+Cesarean	0.010	1,750.53	17.03	0.908	0.0088
Hypertension+Normal delivery	0.047	1,232.87	57.99	0.914	0.0430
NoAC+Cesarean	0.154	1,488.66	228.93	0.923	0.1419
NoAC+Normal delivery	0.744	971.00	722.13	0.930	0.6915
	1.000		1086.55		0.9260

8.3.9 Expected costs and QALY for short term complications over 1 year for offspring

Expected SC care costs and QALY over 1 year for offspring are calculated by folding back the decision tree, similar to that for SC in mothers. The total costs in each arm of TP, FP, FN, and TN were £110.57, £204.02, £107.91 and £107.91, respectively. Expected QALY over one year in offspring are shown in Table 8.18. The results for the expected utility values show a slight difference from each other.

Table 8.18 Expected costs and QALY over one year in offspring

Short term adverse complications	Probability	Costs(£)	Expected cost (£)	Utility	Expected Utitliy
True positive (TP)					
Macrosomia	0.076	623.71	47.70	0.941	0.0720
Shoulder dystocia	0.014	683.48	9.70	0.903	0.0128
Metabolic problems	0.059	896.45	53.17	0.936	0.0555
NoAC	0.850		0.00	1.000	0.8500
	1.000		110.57		0.9903
False negative (FN)					
Macrosomia	0.179	623.71	111.61	0.941	0.1685
Shoulder dystocia	0.035	683.48	23.74	0.903	0.0314
Metabolic problems	0.077	896.45	68.68	0.936	0.0717
NoAC	0.710		0.00	1.000	0.7097
	1.000		204.02		0.9812
False positive (FP)					
Macrosomia	0.097	623.71	60.59	0.941	0.0915
Shoulder dystocia	0.014	683.48	9.24	0.903	0.0122
Metabolic problems	0.042	896.45	38.08	0.936	0.0398
NoAC	0.847		0.00	1.000	0.8469
	1.000		107.91		0.9903
True negative (TN)					
Macrosomia	0.097	623.71	60.59	0.941	0.0915
Shoulder dystocia	0.014	683.48	9.24	0.903	0.0122
Metabolic problems	0.042	896.45	38.08	0.936	0.0398
NoAC	0.847		0.00	1.000	0.8469
	1.000		107.91		0.9903

8.3.10 Conclusion

For SC in mothers, Preeclampsia is the most serious complication for GDM and has the highest utility decrement in pregnancy. Moreover, preeclampsia remains a leading cause of direct maternal deaths in the UK (Milne et al., 2005). Mothers that have had preeclampsia are associated with increased long term risks in later life such as acute renal failure (1-5%), eclampsia (<1%), liver failure or haemorrhage (<1%), and in rare cases stroke and death (NICE, 2011) (Sibai et al., 2005). Therefore, long term complications of preeclampsia would not contribute to the overall QALY loss from death and stroke. In addition, it was conservatively assumed that most infants born with complicaitons do not suffer significant long-term morbidity. However a minority of babies will experience much longer term impairment. For example, shoulder dystocia may result in permanent brachial plexus dysfunction. 4 -16% of shoulder dystocia cases lead to brachial plexus injuries and less than 10% lead to permanent dysfunction (Gherman et al., 1998) (Clements, 2001). It would contribute relatively little to the overall QALY loss from shoulder dystocia. Therefore, long term complications of shoulder dystocia were assumed to not reduce life expectancy of offspring.

Preeclampsia and shoulder dystocia or short term complications in general may not be adequate to present the effectiveness of screening programmes for GDM. With

the low affect of QALY over one year in mothers and offspring, this model will try to examine long term complications in those groups. Long term complications are discussed in the next section.

8.4 Decision analysis of GDM: long term complications

In long term complications (LC), both mothers and offspring remain at high risk of developing type 2 DM in future life (Kim et al., 2002). Three systematic reviews state that women who have had GDM have, at least, about a seven-fold increased risk of developing type 2 DM in the future compared with those without GDM, as mentioned in section 4.5 (Kim et al., 2002) (Kitzmiller et al., 2007) (Bellamy et al., 2009). Moreover, Clausen and colleagues studied glucose tolerance in adult offspring born of women with and without GDM in Denmark. More than 20% of offspring born to mothers with diet-treated GDM presented type 2 DM at the age of 22 years. The aim of screening test for GDM not only reduced risk of short term complications but also reduced risk of developing type 2 DM. Case detection and SC are not the endpoint outcomes for GDM screening tests. Therefore, this section constructed a decision analysis and cost-effectiveness analysis for GDM screening tests to consider both SC (QALY over one year) and LC (lifetime QALY of type 2 DM) together in both mothers and offspring. This model therefore assumes all pregnant women experience their first pregnancy at age 26. In LC, costs and outcomes are taken into account at age 28 till 78 years for mothers and at age 1 till 78 years for offspring. According to this model, costs and utilities were discounted at a baseline rate of 3.5% based on average inflation.

This model assumes that long term consequences are conditionally independent of SC and therefore no link is drawn between SC and LC. This model assumes that all mothers and offspring have a chance to develop type 2 DM in the future. With the effective management of GDM, that includes blood glucose monitoring, oral medication and regulating insulin, it is assumed that there will be a reduced likelihood of developing type 2 DM in both mother and child in LC. Therefore, treatment during gestation is linked to the risk of developing type 2 DM in the TP group. Whereas, for mothers without GDM that receive treatment (FP) and offspring who are born to such mothers, treatment during gestation is not linked to the risk of developing type 2 DM. This section will discuss long term complications in both mothers and offspring. .

8.4.1 Long term complications in mother: decision tree structure

Many guidelines suggest postpartum follow up glucose tolerance tests for mothers with GDM (WHO, 1999a) (ACOG, 2000) (NICE, 2008d) (American Diabetes, 2004) (Hoffman et al., 1998). This model perform long term complication based on the American diabetes Association (ADA) and the Australasian Diabetes in Pregnancy Society (ADIPS), guidelines mentioned in section 4.8, women with a history of GDM should test their blood glucose by FPG or 2h 75g OGTT within 1-3 days post-delivery. If the test is negative, the next test should be delayed until 6-12 weeks after delivery, for women with GDM who do not have DM immediately postpartum. Then, mothers should receive a confirmation test again around their early postpartum visit using 75g OGTT. If glucose levels are normal on all previous follow ups, reassessment of glycaemia should be undertaken at a minimum of 3 year intervals (American Diabetes, 2004) (Hoffman et al., 1998). The construction of the decision tree for LC in mothers in this model is based on the above guidelines. Patients with DM are denoted as DM+. Patients without DM are denoted as DM-. The decision tree which is based on 4 follow-up type 2 DM postpartum screening tests for mothers with a history of GDM begins with two chance nodes for mothers with and without GDM, both in receipt of and not in receipt of treatment. Each chance node within both of the arms present with two possible pathways, DM+ and DM-, as shown in Figure 8.12. For example, during screening in year 1, if the results of the tests are found to be normal, mothers should be re-evaluated 3 further times at 3 yearly intervals. The decision tree terminates when mothers are diagnosed with DM. The 2 h 75 g OGTT method assumed to use in all 4 follow-up type 2 DM in mothers with a history of GDM.

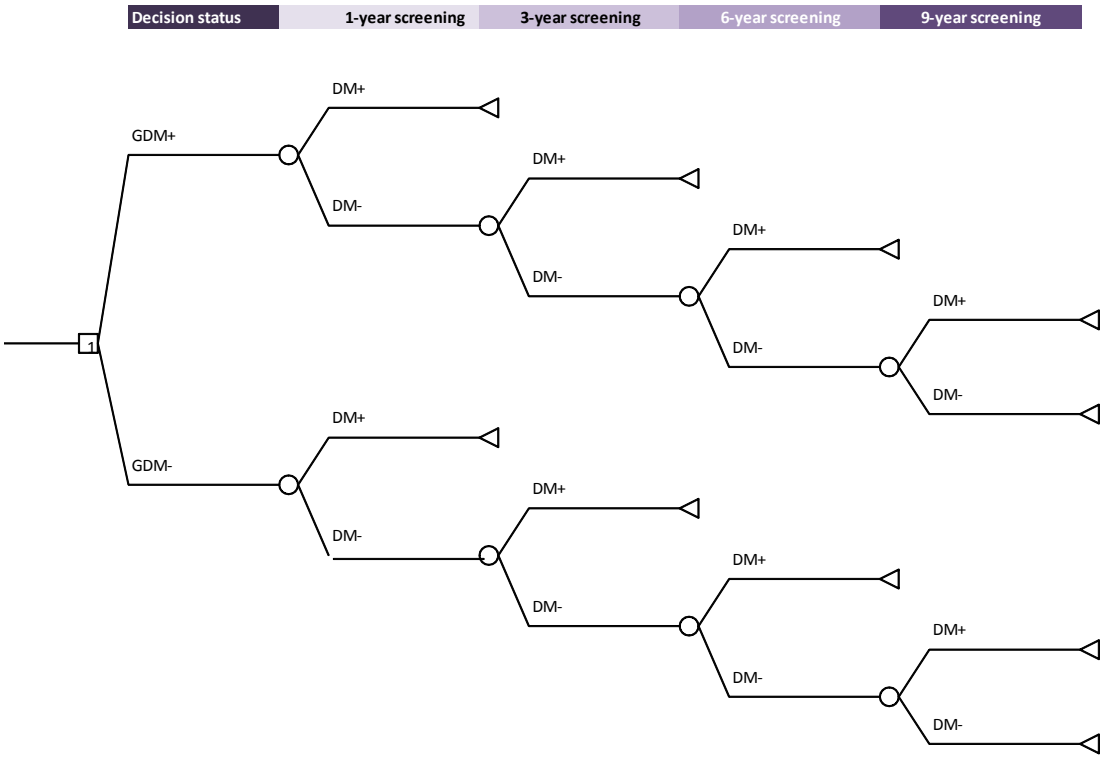


Figure 8.12 Decision tree model for LC in mother

8.4.2 Long term complications in offspring: decision tree structure

There are no guidelines for the screening and treatment of offspring born to mothers with a history of GDM. Therefore, the construction of the decision tree map for offspring is based on the decision tree for mothers. Each chance node within both arms presents with two possible pathways, DM+ and DM-, as shown in Figure 8.13.

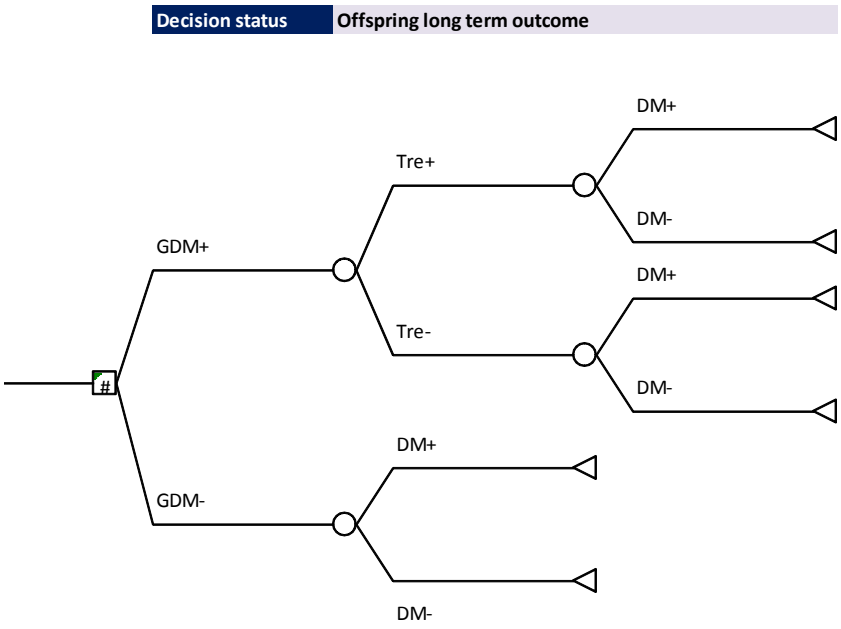


Figure 8.13 Decision tree model for LC in mother

8.4.3 Long term complications: Baseline probabilities

Mothers

In the UK, the average age at first pregnancy is 26 years old (Office for National Statistics, 2013). This model therefore assumes all pregnant women experience their first pregnancy at age 26. The model for the utility for mothers over a 1 year period after delivery was considered from age 27. Therefore, the calculations for utility over a 1 year period after delivery were performed for mothers up until age 28. It is reported that in England and Wales life expectancy at birth for a newborn baby boy is 78.7 years and for a new born baby girl is 82.6, if mortality rates remain the same as they were in 2009-2011 (Office for National Statistics, 2013). In addition, the average Scottish man and women will live to the age of 75.9 and 80.4, respectively (General register office for Scotland., 2011). Therefore, this model assumes life expectancy to be 78 years for all mothers and all surviving offspring, which is the same life expectancy used in the study of cost-effectiveness for GDM and The Office for National Statistic (Werner et al., 2012) (Office for National Statistics, 2013). Therefore, the LC in mothers is calculated from age 28 until age 37. Based on the LC decision tree, mothers are assumed to receive 4 postpartum screening tests at 1, 3, 6 and 9 year postpartum which should be at age 28, 31, 34 and 37, respectively.

In the TP arm, those women who are diagnosed with GDM will have follow up blood glucose tests for the first time at age 28, according to guidelines. Mothers with onset DM at age 28 or with abnormal blood glucose levels at age 28 should start

treatment as a normal DM patient. For mothers with GDM who do not present with DM at 1 year postpartum, reassessment for glycaemia is performed at a minimum of 3 yearly intervals at age 31, 34 and 37 respectively. Therefore the rate of development of Type 2 DM among mothers with and without GDM is required. Assuming no intervention, women with GDM have a 12.6% (95% CI 12.1 -13.9) likelihood of progressing to overt DM (Bellamy et al., 2009). Feig and colleagues used a population-based database to determine the incidence of DM after diagnosis of GDM in Canada. 21,823 women with GDM in the database showed the incidence of DM rising over the 9 year study period. The probability of developing DM was 3.7% at 9 months, 4.9% at 15 months and 13.1 at 5.2 years. By the end of follow up, the incidence was 18.9% at 9 years. In contrast, the incidence of type 2 DM in women without GDM (n= 637,341) was 1.95 at 9 years (Feig et al., 2008) .This model used the figures from the study by Feig and colleagues to present the rate of development of type 2 DM in mothers with GDM. Mothers that have abnormal blood glucose levels in the interval tests are assumed to receive treatment as normal DM patients. Mothers that receive treatment for type 2 DM are assumed to have a lower chance of developing type 2 DM with complications. This model assumes that mothers continue to receive treatment until they die. Mothers with normal blood glucose in each interval test remain in perfect health until diagnose with type 2 DM.

In the other arm, mothers that have a false negative result for GDM do not receive any postpartum screening tests for type 2 DM, do not receive treatment, and may have chance to develop type 2 DM. The same rate of development of type 2 DM, as mentioned in the above paragraph, was used in the decision tree arms for the false negative result for GDM. Moreover, if the mother's blood glucose levels remain higher than normal without treatment over a long period of time, their blood vessels can gradually become damaged; high glucose levels can damage nerves, organs and lead to serious illness. This may lead to a range of complications (often years after first developing diabetes). It is assumed in this model that such mothers may develop type 2 DM with complications; by the time these mothers are diagnosed, those with Type 2 diabetes will show signs of complications. The associated serious complications include heart disease, stroke, blindness, kidney disease, nerve damage and amputations leading to disability and premature mortality.

In the disease status of FP and TN, this model assumed mothers present with type 2 DM without complications. The Scottish diabetes survey 2001 by NHS Scotland reported 247,278 people were diagnosed with diabetes in Scotland, with a prevalence of 4.7% for both type 1 and type 2 DM. The survey reported the incidence rate (per 100,000 populations per year) of type 2 diabetes by age group from 2008 – 2011(NHS

Scotland, 2012). This model applies the incident rate in the report for 2011 to baseline probabilities of type 2 DM in the general population. This is done by converting the incident rate per 100,000 into probabilities.

Offspring

The children of type 2 DM or GDM mothers may be prone to developing type 2 DM in future life, as mentioned in section 4.3.4 (Weintrob et al., 1996). The prevalence of type 2 diabetes in the 20 – 24 year age group was found to vary in different subjects; 1.4% in non-diabetes and 8.6% in offspring of pre-diabetic mothers. In another study, the prevalence of type 2 DM in the adult offspring of women with GDM in the 18 – 27 year age group was 11.7% (Clausen et al., 2008). All mothers with GDM in the Clausen study received diet and lifestyle treatment. This model assumes the age of offspring that develop type 2 DM are between the ages of 20 – 29 years. Additionally the figures in the Clausen study are taken to be representative of probability for the offspring of mothers with GDM.

There are no articles that review the effect of treatment in women with GDM on adult offspring with DM. However, in South Australia a multi-centre randomised control trial looked at the effect of the treatment of mild GDM compared to no treatment on the body mass index (BMI) of 4-5 year olds. The difference between the treatment groups was 1.19 (95% CI 0.78 – 1.82), representing very little change in the BMI for the 4-5 year age group (Gillman et al., 2010). Therefore, this model assumes the relative risk of developing type 2 DM in offspring with GDM that is untreated to be 1.19. In mothers without GDM (FP and TN) the risk of developing DM in the offspring of women without GDM at age 20 – 29 years is used in this arm. This model also applies the Scottish diabetes incident rate in this decision tree (NHS Scotland, 2012).

Considering how the distribution of probability parameters are determined in LC for mothers and offspring, Table 8.19 details the prevalence parameters and distribution used in the probabilistic analysis. Uncertainty in the prevalence of developing type 2 DM in both mothers and offspring were accounted for by fitting a Beta distribution, fitted by using the method of moment to find α and β, given the reported mean value and SE (Briggs et al., 2006).

One of the most common types of parameter used to integrate treatment effects into a model are relative risk parameters. Using the relative risk of developing type 2 DM in adult offspring of 1.19, results in a negative probability. Relative risk values greater than 1 cannot be used and the odds ratio can be employed as an alternative. This is

only a possible consequence of relative risk; it is not found with the odds ratio. The appropriate distribution for the odds ratio is lognormal.

Table 8.19 Long term prevalence parameters in model

Prevalence parameters	Point estimate	SE	Probability distribution	Data source
Mothers with GDM				
1 year postpartum (onset DM at 28y)	0.036	0.001	Beta	(Feig et al., 2008)
3 year postpartum (onset DM at 31y)	0.116	0.002	Beta	(Feig et al., 2008)
6 year postpartum (onset DM at 34y)	0.116	0.001	Beta	(Feig et al., 2008)
9 year postpartum (onset DM at 37y)	0.159	0.002	Beta	(Feig et al., 2008)
Mothers without GDM				
9 year postpartum (onset DM at 37y)	0.020	0.001	Beta	(Feig et al., 2008)
Offspring in Mother with GDM				
Onset DM at 20-29 y	0.214	0.032	Beta	(Clausen et al., 2008)
OR of developing DM with untreated GDM	1.290	0.424	Log normal	(Gillman et al., 2010)
Offspring in mother without GDM				
Age 10-19 years	0.000	0.000	Beta	(NHS Scotland, 2012)
Age 20-29 years	0.002	0.000	Beta	(NHS Scotland, 2012)
Age 30-39 years	0.224	0.001	Beta	(NHS Scotland, 2012)
Age 40-49 years	0.289	0.003	Beta	(NHS Scotland, 2012)
Age 50-59 years	0.144	0.006	Beta	(NHS Scotland, 2012)
Age 60-69 years	0.347	0.008	Beta	(NHS Scotland, 2012)
Age > 70 years	0.726	0.007	Beta	(NHS Scotland, 2012)

8.5 Long term complications: life time cost estimates

Patients with diabetes, over the long term, experience a decrease in quality of life and an increase in the use of health care services (Loukine et al., 2012). The costs for long term complications include postpartum screening for diabetes and treatment costs for type 2 DM with and without complications. For postpartum screening tests, mothers with TP and FP receive postpartum screening for diabetes after 6 weeks, then at an interval of at least once every 1-2 years for women with normal glucose tolerance. The postpartum follow up for type 2 DM is screened for by using a 2 hour 75g OGTT and costs £18.76 a time (NICE, 2008a).

Treatment costs for type 2 DM were obtained from UKPDS 65. The UKPDS 65 estimated costs based on patient-level data from a randomised clinical controlled trial involving 5,102 UKPDS patients, mean age 52.4 years at diagnosis (Clarke et al., 2003). This study was conducted in 23 hospital-based clinics in England, Scotland and Northern Ireland as part of the UK Prospective Diabetes Study (UKPDS). This study presents estimated annual hospital in-patient costs, conditional on some in-patient cost being incurred and the expected mean in-patient costs of complication were found to be

£2,543 (95% CI 2,406-2,697) and £157 (95% CI 145-170), respectively. For out-patients, the estimated costs, conditional on cost being incurred, and the expected mean annualised non-in-patient costs were £70 (95% CI 66-76) and £159 (95% CI 149-173), respectively. This model uses those figures to present annual costs for type 2 DM with and without complications.

The cost parameters and distributions used in the probabilistic analysis are detailed in Table 8.20. The relevant costs were calculated by adjusting the costs in other years to be equivalent to those in 2011, as mentioned in section 7.5.2. As they are fixed costs it is appropriate to hold the costs of screening tests for post partum follow up constant in the probabilistic analysis. A normal distribution is suitable to represent the unit cost parameters in both in-patients and out-patients, and the SE value was calculated from a 95% CI range.

Table 8.20 Long term cost parameters in the model

Cost (annual hospital)	Point estimate (£)	SE	Probability distribution	Data source
Screening test				
postpartum follow up	18.76		Deterministic	(NICE, 2008a)
No complications				
In-patient costs	2,973.00	86.79	Normal	(Clarke et al., 2003)
Out-patient costs	82.57	2.98	Normal	(Clarke et al., 2003)
With complications				
In-patient costs	183.55	7.46	Normal	(Clarke et al., 2003)
Out-patient costs	187.75	4.18	Normal	(Clarke et al., 2003)

Life time costs for mother and offspring with diabetes are shown in Table 8.21. Life time costs of type 2 DM without complications are about 1.1 times lower than life time costs for type 2 DM with complications. Life time cost of type 2 DM with an onset at age 20 – 29 presented with the highest value at £ 170,518.85.

Table 8.21 Life time costs for mothers and offspring by disease status

Age at diagnosis	Disease status	Cost (£)
Mother		
1 year postpatum (onset at 28y)	DM +	147,397.65
	DM + with C	165,523.56
3 year postpatum (onset at 31y)	DM +	138,727.20
	DM + with C	155,786.88
6 year postpatum (onset at 34y)	DM +	130,056.75
	DM + with C	146,050.20
9 year postpatum (onset at 37y)	DM +	121,386.30
	DM + with C	136,313.52
Offspring		
On set at 20-29	DM +	170,518.85

8.6 Life table and life expectancy

This model estimates the life time health outcomes for mothers and offspring with and without type 2 DM and calculates outcomes in terms of mean life expectancy (LE), defined as the average number of years a person life expectancy is. Moreover, they can also be used to estimate future health care costs for patients with type 2 DM and quality-adjusted life expectancy, which will be mentioned in the next section. This calculation starts with life tables for both mothers and offspring at different ages. Complete life tables are a convenient way to analyse age specific death rates and are a standard demographic tool which are used to outline life expectations for a range of ages (Wesley, 1998). This model created life tables for mothers and offspring based on mortality rate data from the Office for National Statistic in order to estimate life expectancy (Office for National Statistics, 2013). Life tables for mothers and offspring with perfect health start at age 27 and age 1, respectively.

Life table analysis was applied to calculate the life expectancy impact of DM. To calculate the life expectancy for Patients with type 2 DM (DM+), mortality rates from the Office for National Statistics are multiplied by relative differences in death rates to diabetes sourced from population-based cohorts of 4,842 people with diabetes living in South Tees in the UK (Roper et al., 2002). This study reported the death rate per 1,000 person-years of type 2 DM at ages 40-59, 60-79 and 80+ as 22.14, 57.12 and 155.56, respectively, and relative risk (RR) to the all-cause mortality rate of type 2 DM at ages 40-59, 60-79 and 80+ were 3.30 (95% CI 2.15-3.95), 1.41 (95% CI 1.28-1.56) and 1.09 (95% CI 0.92-1.29), respectively. The death rate reported in this study was not used because death rates per 1000 are particular to the population studied and are likely to differ from this particular population. Relative rates are generally thought to be more generalisable and give a better estimate of the underlying/baseline rate from UK life tables.

One of the most common types of parameter used to integrate life tables into a model are relative risk parameters. The appropriate distribution for the RR is lognormal. The various RR used in probabilistic analysis are detailed in Table 8.22.

Table 8.22 DM parameters in model

Prevalence parameters	Point estimate	SE	Probability distribution	Data source
Death with DM				
Age 30-49 years	3.15	0.116	Log normal	(Roper et al., 2002)
Age 50-64 years	1.41	0.344	Log normal	(Roper et al., 2002)
Age 65-74 years	1.09	0.086	Log normal	(Roper et al., 2002)

8.7 Long term complications: quality-adjusted life expectancy

UK Population Norms for EQ-5D were used to calculate quality-adjusted life expectancy in healthy mothers and offspring. The EQ-5D result is nationally representative of 3,395 men and women aged 18 or over living in the UK interviewed for the survey (Paul et al., 1999). The mean weighted health state index by age and standard region (Scotland) are used to calculate the quality-adjusted life expectancy in healthy offspring. Moreover, the mean weighted health state index by age and standard region (Scotland) for females is used to calculate the quality-adjusted life expectancy in healthy mothers. Life expectancy for each age in the life tables were multiplied by the mean utility in each age group of EQ-5D to present quality-adjusted life expectancy in healthy mothers and offspring. Mothers with GDM may be concerned about how their babies will be affected in the future, and this could have an additional impact on the utility of the mothers. However, in this model, the utility of mothers with GDM is calculated from the health related quality of life for short and long term complications. Consequently, the disutility of mothers with relation to their offspring was ignored.

The utilities from UK population Norms for EQ-5D are suited to a Beta distribution, by using the method of moment to fit Beta distribution to the utility data. Table 8.23 details the utility and distributions used in probabilistic analysis in terms of calculating quality-adjusted life expectancy in a normal population.

Table 8.23 UK population Norms for EQ-5D in all sexes and females in Scotland

Utility	Point estimate	SE	Probability distribution	Data source
UK Population Norms for EQ-5D (all sexes)				
Under 25 year	0.950	0.015	Beta	(Paul et al., 1999)
Age 25 – 34 years	0.910	0.021	Beta	(Paul et al., 1999)
Age 35 – 44 years	0.920	0.019	Beta	(Paul et al., 1999)
Age 45 – 54 years	0.880	0.035	Beta	(Paul et al., 1999)
Age 55 – 64 years	0.820	0.027	Beta	(Paul et al., 1999)
Age 65 – 74 years	0.730	0.043	Beta	(Paul et al., 1999)
Over 75 years	0.7840	0.045	Beta	(Paul et al., 1999)
UK Population Norms for EQ-5D (female)				
Under 25 year	0.950	0.024	Beta	(Paul et al., 1999)
Age 25 – 34 years	0.890	0.031	Beta	(Paul et al., 1999)
Age 35 – 44 years	0.900	0.029	Beta	(Paul et al., 1999)
Age 45 – 54 years	0.910	0.025	Beta	(Paul et al., 1999)
Age 55 – 64 years	0.810	0.041	Beta	(Paul et al., 1999)
Age 65 – 74 years	0.680	0.064	Beta	(Paul et al., 1999)
Over 75 years	0.750	0.059	Beta	(Paul et al., 1999)

8.7.1 Quality-adjusted life expectancy with DM

DM occurs as a result of GDM, therefore, quality of life associated with diabetes mellitus was assumed to represent the QALY in LC for GDM. In fact, individuals with diabetes have reduced HRQol compared with those without diabetes in the same age group with diabetes-related complications (Solli et al., 2010). The effect of DM on Quality of life has been addressed by a number of authors (Diabetes Control and Complications Trial Research Group, 1996) (Testa et al., 1998) (Hanninen et al., 1998) (Wu et al., 1998) . Various tools that have been employed to estimate quality of life in diabetes including SF-36 (Ware and Sherbourne, 1992) (Wu et al., 1998), SF-20 (Hanninen et al., 1998), EQ-5D (Solli et al., 2010) and others (Testa et al., 1998) (Brown et al., 2000). In Norway, the study in HRQol in diabetes mellitus was conducted on 1,000 individuals with type 1 and type 2 diabetes using EQ-5D (Solli et al., 2010). All patients with DM completed EQ-5D questionnaires by using a time trade-off method. For DM without complications, the EQ-5D index for type 2 DM was 0.85 (95%CI 0.82-0.87). There was another study in the Netherlands which aimed to estimate the HRQoL and treatment satisfaction for patients with type 2 DM (Redekop et al., 2002). Of the 1,348 Dutch type 2 DM patients surveyed, the EQ-5D utility scores reported by age groups <50 years, 50-59 years, 60-69 years and equal and over 70 years were 0.79 (SD 0.26), 0.75 (SD 0.26), 0.78 (SD 0.25) and 0.70 (SD 0.28), respectively. To calculate the quality adjusted life expectancy with DM, this model used all utility figures from the Netherlands study and multiplied them by the life expectancy in each age group. In order to determine QALY after type 2 DM develops, this model assumes all patients who received screening and treatment do not suffer from any complications of DM until the end of their life at age 78.

The HRQol in type 2 DM is appropriate for a Beta distribution, by using the method of moment to fit Beta distribution to the utility data. Table 8.24 outlines the utility and distributions used in probabilistic analysis with respect to the calculation of quality-adjusted life expectancy in patients with type 2 DM.

Table 8.24 HRQol in type 2 DM

Utility	Point estimate	SE	Probability distribution	Data source
HRQoL in type 2 DM				
Under 50 year	0.790	0.008	Beta	(Redekop et al., 2002)
Age 50 - 59 years	0.750	0.008	Beta	(Redekop et al., 2002)
Age 60 – 69 years	0.780	0.007	Beta	(Redekop et al., 2002)
Over 75 years	0.700	0.008	Beta	(Redekop et al., 2002)

8.7.2 Quality-adjusted life expectancy with DM with complications

Diabetes without treatment results in high glucose levels in the blood, which can damage nerves, organs and lead to serious illness. Diabetes with complications leads to poor health, premature mortality, and to a reduction of life expectancy and quality-adjusted life expectancy (Loukine et al., 2012).

In the Solli and colleagues study, the EQ-5D dimensions were able to capture diabetes related complications with the strongest determinants of reduced HRQoL as being ischemic heart disease, stroke and neuropathy. For type 2 DM with complications the EQ-5D index was 0.73 (95% CI 0.69-0.78) and for more than 2 complications it was 0.64 (95%CI 0.56-0.71) (Solli et al., 2010).

The two figures of EQ-5D in diabetes with complications in this study are used to calculate life time QALY of DM with complications. This model assumes that patients develop type 2 DM and that when they receive a diagnosis, those with Type 2 DM will present with a single complication. Moreover, this model assumes that twenty years after diagnoses patients will present with more than 2 complications.

The HRQol in type 2 DM with complications suits a Beta distribution, by using the method of moment to fit Beta distribution to the utility data. Table 8.25 details the utilities and distributions used in probabilistic analysis in terms of calculating quality-adjusted life expectancy in patients presenting with type 2 DM with complications.

Table 8.25 HRQol in type 2 DM with complication

Utility	Point estimate	SE	Probability distribution	Data source
HRQoL in type 2 DM with complication				
Type 2 DM with any complication	0.730	0.082	Beta	(Solli et al., 2010)
Type 2 DM with >2 complications	0.640	0.038	Beta	(Solli et al., 2010)

Table 8.26 presents the results of life expectancy for mothers with DM- and DM+. The total life years of mothers at age 27 with perfect health and for those suffering from DM with no complications were 74 (27+47.79) and 72 (27+45.85), respectively. The life expectancies for DM in this model were slightly high. The low mortality rate of diabetes identified in the paper by Roper and colleagues (2002) may influence the high life expectancy in mothers with DM.

Table 8.26 Estimates of life expectancy and Quality-adjusted life expectancy in mothers by age at diagnosis of type 2 DM

Mother		Life time	Life year lost	QALY	QALY lost
1 year postpatum (onset at 28y)	DM-	47.79		39.84	
	DM +	45.85	1.94	35.17	4.67
	DM + with C			31.95	7.89
3 year postpatum (onset at 31y)	DM-	44.84		37.21	
	DM +	43.00	1.84	32.92	4.29
	DM + with C			29.86	7.35
6 year postpatum (onset at 34y)	DM-	41.90		34.59	
	DM +	40.18	1.72	30.69	3.90
	DM + with C			27.80	6.79
9 year postpatum (onset at 37y)	DM-	38.97		31.96	
	DM +	37.41	1.56	28.49	3.47
	DM + with C			25.76	6.20

However, life expectancy based on these calculations gives a similar result as the United Kingdom Prospective Diabetes Study (UKPDS) outcome model (UKPDS no.68). The UKPDS developed a simulation model for type 2 DM that can be used to estimate the likely occurrence of major diabetes-related complications over a lifetime, in order to calculate life expectancy and quality adjusted life year (Clarke et al., 2004). Table 8.27 shows a comparison of the estimate of life expectancy and quality-adjusted life expectancy with DM by age at diagnosis of type 2 DM, and as can be seen, the results for estimated life expectancy in mothers show similar results to the UKPDS study (Clarke et al., 2004).

Table 8.27 Estimations of life expectancy with DM+ by age at diagnosis of type 2 DM

Age	Life expectancy		Qauality-adjusted life expectancy	
	DM+	DM+ (UKPDS)	DM+	DM+ (UKPDS)
45-49	30.27	26.80*	22.82	20.26*
50-54	25.95	23.58*	19.38	17.62*
55-59	22.32	20.16*	14.64	15.10*
60-64	17.02	17.65*	12.68	13.15*

*Based on Table 5 in study UKPDS no.68

The same method is used to calculate life expectancy in offspring; life expectancies of offspring with and without DM in the 20-29 age groups were 74 (20+ 54.82) and 72 (20+52.41) years, respectively, as show in Table 8.28. This estimate is slightly similar to the UKPDS study. For the 50-54 age groups, the UKPDS outcome model reported life expectancy and quality-adjusted life expectancy with DM to be 23.58 and 17.62, respectively. Again this estimate is similar to the results for the 50-59 age groups. Life expectancy and quality-adjusted life expectancy with DM are 25.94 and 25.26, respectively.

Table 8.28 Estimate of life expectancy and Quality-adjusted life expectancy with DM in offspring by age at diagnosis of type 2 DM

Offspring		Life time	Life year lost	QALY	QALY lost
Onset at 10-19 y	DM-	64.70		56.53	
	DM +	62.29	2.41	48.23	8.31
Onset at 20-29 y)	DM-	54.82		47.14	
	DM +	52.41	2.41	40.42	6.72
Onset at 30-39 y	DM-	45.04		37.99	
	DM +	42.67	2.37	32.72	5.27
Onset at 40-49 y	DM-	35.38		29.12	
	DM +	33.67	1.70	25.58	3.53
Onset at 50-59 y	DM-	25.94		20.57	
	DM +	25.26	0.68	18.88	1.69
Onset at 60-69 y	DM-	16.88		12.80	
	DM +	16.56	0.32	12.35	0.45
Onset at >70 y	DM-	8.18		6.01	
	DM +	8.11	0.07	5.68	0.33

8.7.3 Expected costs and QALY for long term complications in mothers

The probabilities in section 8.4.3 were assigned to each of the outcomes in the decision tree branches. Then, the expected costs and quality-adjusted life expectancy of LC in offspring were calculated by multiplying the substituted actual numerical variables in each branch with the monetary cost and QALYs. This model uses the folding back technique to calculate costs and QALYs in TP, FN, FP, and TN, as shown in Table 8.29. The highest expected life time costs for LC were £57,073.81 in FN. The lowest expected quality-adjusted life expectancy was 30.57 in FN. However, the results of expected QALYs should be ordered from the worst outcomes to the best, namely FN, TP, FP and TN, respectively. For the expected QALYs for long term complications, the FP (31.96) arm appears to be associated with greater expected QALYs than TP (31.64). As mentioned at the end of section 8.4, long term consequences are conditionally independent of SC and therefore no link is drawn between SC and LC. For mothers in the FP groups, disutility is taken into account for SC. Therefore, the FP group appears to be lower than TP for SC. With the low side effects of treatment during gestation, the effects will disappear after delivery. Mothers with FP tests have a chance to receive postpartum screening which will allow the early detection of type 2 DM. Moreover, mothers with FP from SC have an opportunity to receive post partum screening to prevent type 2 DM with complications. Mothers with FP from SC therefore receive additional benefit from this screening with respect to LC. Therefore, the FP arm presents with greater expected QALYs than the TP arm in LC.

Table 8.29 Expected lifetime costs and QALY in mothers

Age onset DM after postpartum	Probabiltiy	Life time cost (£)	Expected costs(£)	life time QALY	Expected QALY
True positive (TP)					
Screening at 1 year (onset at 28y) DM+	0.0357	155,834.07	5,557.14	35.17	1.25
Screening at 3 year (onset at 31y) DM+	0.1117	146,686.12	16,380.17	32.92	3.68
Screening at 6 year(onset at 34y) DM+	0.0987	137,538.18	13,580.14	30.69	3.03
Screening at 9 year (onset at 37y) DM+	0.1198	128,390.23	15,384.98	28.49	3.41
Screening at 9 year (onset at 37y) DM-	0.6341	75.06	47.60	31.96	20.27
	1.0000	568,523.67	50,950.02		31.64
False negative (FN)					
NoScreening at 1 year (onset at 28y) DM+	0.0357	174,770.37	6,232.42	31.95	1.14
NoScreening at 3 year (onset at 31y) DM+	0.1117	164,489.76	18,368.27	29.86	3.33
NoScreening at 6 year (onset at 34y) DM+	0.0987	154,209.15	15,226.19	27.80	2.74
NoScreening at 9 year(onset at 37y) DM+	0.1198	143,928.54	17,246.93	25.76	3.09
NoScreening at 9 year(onset at 37y) DM-	0.6341	0.0000	0.00	31.96	20.27
	1.0000	637,397.82	57,073.81		30.57
False positive (FP)					
Screening at 1 year (onset at 28y) DM+	0.0002	155,834.07	35.84	35.17	0.01
Screening at 3 year (onset at 31y) DM+	0.0011	146,667.36	165.60	32.92	0.04
Screening at 6 year(onset at 34y) DM+	0.0011	137,500.65	155.08	30.69	0.03
Screening at 9 year (onset at 37y) DM+	0.0011	128,333.94	144.58	28.49	0.03
Screening at 9 year (onset at 37y) DM-	0.9964	75.06	47.60	31.96	31.85
	1.0000	568,411.08	548.69		31.96
True positive (TN)					
NoScreening at 1 year (onset at 28y) DM+	0.0002	155,834.07	35.84	35.17	0.01
NoScreening at 3 year (onset at 31y) DM+	0.0011	146,667.36	165.60	32.92	0.04
NoScreening at 6 year (onset at 34y) DM+	0.0011	137,500.65	155.08	30.69	0.03
NoScreening at 9 year(onset at 37y) DM+	0.0011	128,333.94	144.58	28.49	0.03
NoScreening at 9 year(onset at 37y) DM-	0.9964	0.00	0.00	31.96	31.85
	1.0000	568,336.02	501.09		31.96

8.7.4 Expected costs and QALY for long term complications in offspring

Based on the probabilities in section 7.4.3, this model assigns a probability value to each outcome branch. The sum of the probabilities of all outcome branches from a single chance node must equal one. The substituted actual numerical variables in each branch were multiplied by the monetary costs and QALYs. The sum of those processes is called the expected value of the node. This model folded back the LC decision tree to present costs and QALYs in TP, FN, FP, and TN, as shown in table 7.22. Table 8.30 shows the expected life time cost and quality-adjusted life expectancy in offspring. The expected cost in FN was the highest costs at £26,366. Whereas, offspring born with GDM mothers without treatment showed the lowest life time QALY at 46.16. Expected QALYs for FP is 47.14, a value that is greater than the expected QALYs for TP (46.35). This is also the case for TP and FP outcomes for expected QALYs in mothers in the previous section. Children with GDM mothers may be more prone to developing type 2 DM in future life than children born to mothers without GDM, for the same reason that offspring born to FP mothers have a lower risk of developing type 2 DM than offspring

born to TP mothers. Therefore, offspring born to FP mothers appear to be associated with greater expected QALYs than those born to TP mothers.

Table 8.30 Expected life time costs and QALY in offspring

Long term complications	Probability	Costs(£)	Expected cost (£)	Utility	Expected Utitliy
True positive (TP)					
DM+	0.1173	180278.63	21,140.17	40.42	4.74
DM-	0.8827			47.14	41.61
	1.0000		21,140.17		46.35
False negative (FN)					
DM+	0.1463	180278.63	26,366.00	40.42	5.91
DM-	0.8537			47.14	40.24
	1.0000		26,366.00		46.16
False positive (FP)					
DM+	0.0002	180278.63	41.46	40.42	0.0093
DM-	0.9998			47.14	47.13
	1.0000				47.14
True negative (TP)					
DM+	0.0002	180278.63	41.46	40.42	0.0093
DM-	0.9998			47.14	47.13
	1.0000				47.14

8.8 Conclusion

The process of populating the outcomes of screening tests for GDM models involves bringing together all the relevant evidence and synthesizing it appropriately in terms of input parameters within the model. This study defines SC mothers as hypertensive, and as having preeclampsia and caesarean sections and the SC for offspring include macrosomia, metabolic problems and shoulder dystocia. The SC show very low impact on QALYs in mothers with GDM. Therefore, this economic evaluation of screening tests for GDM takes into account the LC in both mothers and offspring to show the large impact of screening programs. This cost-effectiveness study of screening tests for GDM also presented the results in four different outcomes. Mothers with FN showed significant high treatment costs and low QALYs when compared to other outcomes. Therefore, in the negative dominated strategy there is a probability of false negative case identification and the likelihood of higher costs and expenditure for the treatment of SC and LC. On the other hand, the PDS showed a probability of FN, while the cost of treating false negatives FP was not high and QALY was similar to that of healthy pregnancy. However, the results from chapter 7 were based on a deterministic model. The next chapter will present the results of cost-effectiveness analysis based on a probabilistic model, which might give different results.

Chapter 9 A decision analysis for the cost-effectiveness analysis of screening tests for gestational diabetes mellitus

9.1 Introduction

In the previous chapter, having detailed the development of the models for screening tests for GDM and how they were parameterised for both mothers and offspring, this chapter presents the deterministic and probabilistic results. In the models, the effectiveness of screening tests for GDM were measured in quality adjusted life years (QALY) with regard to short and long term complications. The beginning of this chapter details the deterministic results for each model reporting the incremental cost per QALY gain in short term complications and life time QALY in long term complications. In the subsequent section, an exploration of uncertainty through probabilistic sensitivity analysis and value of information analysis is reported.

9.2 Deterministic results

The models of screening tests for GDM were developed in order to estimate costs and outcomes in both mothers and offspring of four alternative strategies versus a strategy of “no screening”. In this economic evaluation, decision trees are designed to assess the five strategies of GDM and were built based on the combinations of test approaches that include the negative dominant strategy (NDS) and the positive dominant strategy (PDS). Eight strategies were employed for a comparison of cost-effectiveness analysis, as mentioned in section 7.7.2. The evaluations were undertaken from a health care provider’s perspective (NHS) to inform policy makers on this important issue by establishing the optimal cost-effective screening and diagnostic tests of GDM. This section presents the deterministic results in both short and long term complications.

9.2.1 Short term complications

The costs and consequences in terms of QALY of short term complications from both mothers and offspring are added up to present the cost-effectiveness of GDM screening tests for short term complications. In the mother, the expected costs include screening test costs, treatment care costs during gestation period, and treatment care costs for

adverse complications during gestation delivery. In offspring, the only expected costs taken into account are treatment care costs for adverse complications.

The total cost and QALYs generated for each strategy under the baseline assumptions are presented in Table 9.1. In the mother, costs and QALY in short term complications are presented as C1 and Q1. Similarly, C2 and Q2 are costs and QALY of short term complications in offspring. Costs in both short (C1) and long (C2) term complications are added up to present the total cost of screening for GDM in each strategy. Similarly, QALY gained for screening tests of GDM include both mothers (Q1) and offspring (Q2) term. S5 has the lowest cost (£1,202), whereas the highest QALY gain is for S3 (1.9162) for screening tests of GDM in short term complications.

Table 9.1 Summaries of short term complications in mothers and offspring

Strategy	Short term adverse complications					
	Mothers		Offspring		Total	
	Costs(C1)	QALY(Q1)	Costs(C2)	QALY(Q2)	Costs (C1+C2)	QALY (Q1+Q2)
S1 (NDS)	1,108	0.9259	109	0.9902	1,217	1.9161
S2 (NDS)	1,106	0.9259	111	0.9900	1,216	1.9159
S3 (NDS)	1,112	0.9259	108	0.9903	1,220	1.9162
S4 (NDS)	1,109	0.9259	110	0.9901	1,219	1.9160
S5 (NDS)	1,091	0.9259	111	0.9900	1,202	1.9158
S1 (PDS)	1,158	0.9247	108	0.9903	1,266	1.9150
S2 (PDS)	1,165	0.9247	108	0.9903	1,273	1.9150
S3 (PDS)	1,112	0.9259	108	0.9903	1,220	1.9162
S4 (PDS)	1,147	0.9252	108	0.9903	1,255	1.9155
S5 (PDS)	1,091	0.9259	111	0.9900	1,202	1.9158

Table 9.2 presents the analysis of the incremental cost-effectiveness ratios (ICER) for the analysis of screening tests for GDM in short term complications. The ICER examines the additional costs that one strategy incurs over another and compares this with the additional benefits. The alternative screening test strategies are ranked according to their costs. S5 has the lowest cost but not the highest QALY gain when compared with the other strategies. Next, S2 (NDS) and S4 (NDS) are seen to be more costly and have less QALY gained than S1 (NDS). Therefore, S2 (NDS) and S4 (NDS) are dominated by S1 (NDS). S3, which is less expensive and highly effective, dominates over S1 (PDS), S4 (PDS) and S2 (PDS). The screening test strategies that are the most expensive and less effective are then excluded and the ICER recalculated again.

Table 9.2 Cost per QALY gain in screening tests for GDM in short term complications in mother and offspring

Strategy	Costs (£) [C]	Effects [E]	Incremental costs [△C]	Incremental effect [△E]	ICER [△C/△E]	
S5	1,202	1.9158	0.00	0.0000	0	
S1 (NDS)	1,217	1.9161	14.47	0.0003	46760	
S2 (NDS)	~~1,216~~	~~1.9159~~	~~-0.56~~	~~-0.0002~~	~~2515~~	Dominate
S4 (NDS)	~~1,219~~	~~1.9160~~	~~2.56~~	~~0.0001~~	~~27250~~	Dominate
S3	1,220	1.9162	0.92	0.0002	4282	
S1 (PDS)	~~1,266~~	~~1.9150~~	~~46.36~~	~~-0.0012~~	~~0~~	Dominate
S4 (PDS)	~~1,255~~	~~1.9155~~	~~-11.37~~	~~0.0005~~	~~0~~	Dominate
S2 (PDS)	~~1,273~~	~~1.9150~~	~~18.75~~	~~-0.0005~~	~~-38914~~	Dominate

The ICER was recalculated for S5, S1 (NDS) and S3 after excluding the dominated screening strategies, S2 (NDS), S4 (NDS), S1 (PDS), S4 (PDS) and S2 (PDS), as shown in Table 9.3.

Table 9.3 Exclusion of more costly and less effective alternatives

Strategy	Costs (£) [C]	Effects [E]	Incremental costs [△C]	Incremental effect [△E]	ICER [△C/△E]
S5	1,202	1.9158	0	0.0000	0
S1 (NDS)	1,217	1.9161	14	0.0003	46760
S3	1,220	1.9162	3	0.0001	33969

S5 is the “Do Nothing” strategy and represents the comparator in this model. S5 is less costly than the other screening strategies as there are no screening costs associated with the strategy. Additionally as S5 involves no screening the only possible results for both mothers and offspring are false negative (FN) and true negative (TN).

The next most effective strategies are S1 (NDS) and S3 when costs and effectiveness are compared with S5. S1 (NDS) and S3 are universal screening strategies developed by SIGN 2001 and Consensus 2010, respectively (SIGN, 2010) (Moses, 2010). S1 (NDS) is therefore the second most cost-effective screening strategy for short term complications in mothers and offspring, as it was for case identification, as mentioned in sections 7.7.2.

Figure 9.1 presents the CE plane for short term complications, with an x-axis of QALY over one year and y-axis of costs in pounds sterling. The total costs and total effectiveness for each strategy are plotted on the CE plane. On this plane, the screening test strategies are plotted against S5, shown as the origin of the graph. As can be seen from the graph, for S5 the costs and QALY gained per year were £1,202 and 1.9158 respectively. Extended dominance is also illustrated in the figure; S4 (NDS) is

dominated by S2 (NDS) and S1 (NDS). If S4 (NDS) is used to screen all patients in the population, this will cost £1,219 and the total effectiveness will be 1.9160. However, 1.9160 effectiveness units can be achieved at a low cost by using a combination of S2 (NDS) and S1 (NDS).

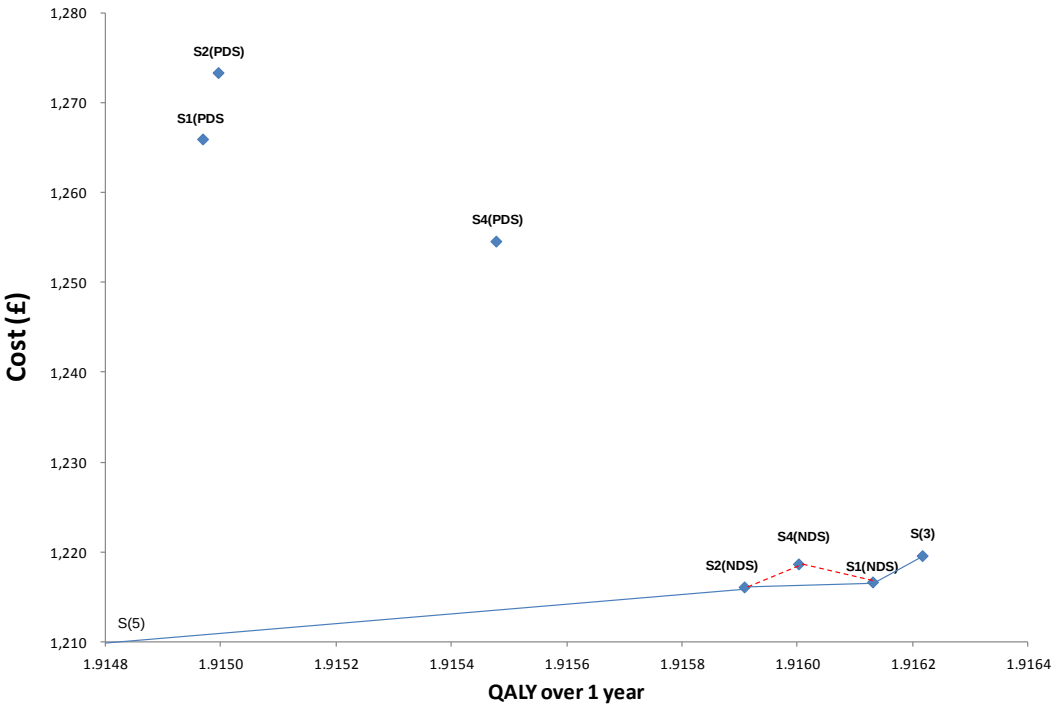


Figure 9.1 the cost-effectiveness plane for short term complications

9.2.2 Long term complications in mothers and offspring

In the previous section, the cost-effectiveness analysis is presented as short term clinical outcomes for both mothers and offspring in terms of costs and QALY over 1 year. An economic evaluation requires the analytical timeframe to cover a period of time that takes into consideration the final outcome relevant to the decision being made. In this section, the costs and QALY for short term and long term complications in mothers and offspring are added up to represent long term complications in GDM. The total costs included the cost of treatment during gestation, short term complication treatment costs and lifetime treatment costs for type 2 DM, while the utility included QALY over one yere and health-related quality of life for type 2 DM in mothers and offspring.

With the same GDM prevalence of 3.5% as before, the long term costs and consequences in both mothers and offspring are presented in Table 9.4 in terms of cost and effectiveness. Costs in both mothers and offspring are presented as C1, C2, C3 and C4. Similarly, QALY of both mothers and offspring are present as Q1, Q2, Q3 and Q4.

All costs and QALY are added up to present the total cost of long term complications for GDM in each strategy. For short term complications, NDS is less expensive and more effective than PDS, as mentioned in the previous section. However for long term complications, PDS is less expensive and more effective than NDS. After adding up the total costs and QALY for both short and long term complications, PDS becomes the dominant screening test strategy. For long term complications, the strategy with the lowest costs and highest effectiveness was S4 (PDS) at £4,088 and 80.9736, while the strategy with the highest costs and lowest effectiveness was S5 at £4,620 and 80.9287, respectively.

Table 9.4 Summaries of long term compilations in mothers and offspring

Strategy	Short term complications				Long term complications				Total	
	Mother		Offspring		Mother		Offspring		Cost	QALY
	Cost(C1)	QALY(Q1)	Cost(C2)	QALY(Q2)	Cost(C3)	QALY(Q3)	Cost(C4)	QALY(Q4)	(C1+C2+C3+C4)	(Q1+Q2+Q3+Q4)
S1 (NDS)	1,086	0.9261	109	0.9903	2,335	31.9338	842	47.1069	4,372	80.9571
S2 (NDS)	1,091	0.9261	110	0.9902	2,396	31.9232	894	47.1050	4,491	80.9445
S3 (NDS)	1,097	0.9262	108	0.9904	2,259	31.9472	776	47.1094	4,240	80.9731
S4 (NDS)	1,095	0.9261	110	0.9902	2,375	31.9269	876	47.1057	4,455	80.9489
S5 (NDS)	1,078	0.9261	111	0.9901	2,472	31.9100	959	47.1026	4,620	80.9287
S1 (PDS)	1,110	0.9257	108	0.9904	2,273	31.9472	776	47.1094	4,268	80.9727
S2 (PDS)	1,136	0.9249	108	0.9903	2,315	31.9472	776	47.1094	4,335	80.9718
S3 (PDS)	1,097	0.9262	108	0.9904	2,259	31.9472	776	47.1094	4,240	80.9731
S4 (PDS)	1,124	0.9253	108	0.9903	2,079	31.9485	776	47.1094	4,088	80.9736
S5 (PDS)	1,078	0.9261	111	0.9901	2,472	31.9100	959	47.1026	4,620	80.9287

The analysis of the incremental cost-effectiveness ratios (ICER) of the different screening tests for long term complications of GDM are shown in Table 9.5. The different screening test strategies are ordered based on costs. S4 (PDS) has higher QALY (80.9736) and is less expensive (£4,088) than the other strategies. S4 dominates the other screening test strategies for long term complications.

Table 9.5 Costs per life time QALY gained in screening tests for GDM

Strategy	Costs (£) [C]	Effects [E]	Incremental costs [△C]	Incremental effect [△E]	ICER [△C/△E]	
S4 (PDS)	4,088	80.9736				
S3	~~4,240~~	~~80.9731~~	~~152.63~~	~~-0.0004~~	~~-345,648~~	Dominated
S1 (PDS)	~~4,268~~	~~80.9727~~	~~27.31~~	~~-0.0005~~	~~-59,343~~	Dominated
S2 (PDS)	~~4,335~~	~~80.9718~~	~~67.40~~	~~-0.0008~~	~~-80,602~~	Dominated
S1 (NDS)	~~4,372~~	~~80.9571~~	~~37.43~~	~~-0.0147~~	~~-2,547~~	Dominated
S4 (NDS)	~~4,455~~	~~80.9489~~	~~82.47~~	~~-0.0083~~	~~-9,986~~	Dominated
S2 (NDS)	~~4,491~~	~~80.9445~~	~~36.13~~	~~-0.0044~~	~~-8,176~~	Dominated
S5 (PDS)	~~4,620~~	~~80.9287~~	~~128.44~~	~~-0.0157~~	~~-8,156~~	Dominated

The CE plane for long term complications, with an x-axis of lifetime QALY for type 2 DM and a y-axis of costs in pounds sterling is presented in Figure 9.2. On the CE plane, the screening test strategies are plotted against S5, shown as the origin of the graph. As can be seen from the graph, for S4 the costs and QALY gained per year were £4,088 and 80.9736 respectively.

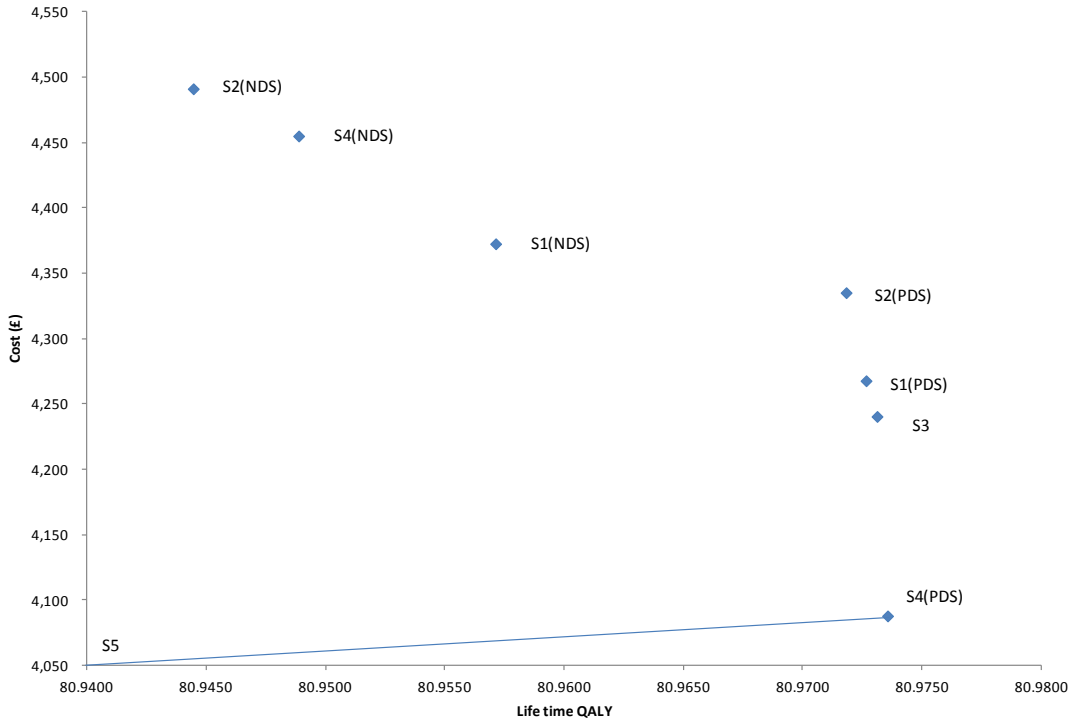


Figure 9.2 the cost-effectiveness plane for long term complications

S4 is a negative dominant selective screening strategy developed by SIGN 2010 (SIGN, 2010), whereby all pregnant women with one or more high risk factors are requested to undertake 1h and 2h 75g oral glucose tolerance test (OGTT) diagnostic tests. On the other hand, S4 (PDS) involves the screening of all pregnant women with one or more risk factors and if the mothers test positive for risk factors they then receive treatment. However, mothers that do not present with risk factors undergo a diagnostic test with 75g OGTT. If the diagnostic test results are over a specified cut-off point, mothers receive treatment. This selective screening strategy shows more cost effectiveness than all of the other screening strategies for long term complications. Regarding the comparison of the test approach combinations, PDS is dominant to NDS in long term complications, with PDS being able to prevent both mothers and offspring developing type 2 DM with complications, as shown in Table 9.5.

S3, however, is a one step approach using 75g OGTT as a gold standard with 100% sensitivity and 100% specificity, with the probability of true positive (TP) and true

negative (TN) results being equal to one. Therefore, S4 (PDS) and S3 give the same probability of TP at 3.5%. The reason that the results of the total costs of S4 (PDS) are lower than S3 for long term complications is due to the cost of the screening tests themselves; risk factor screening tests (£9.50) are less expensive than diagnostic tests (£18.57).

9.3 Probabilistic results

The probabilistic analysis explores uncertainty in cost-effectiveness outcomes. The two economic models of screening tests for GDM in both mothers and offspring are analysed probabilistically using Monte Carlo simulation (1000 iterations) in order to determine expected costs, outcomes and cost-effectiveness. The costs relate to lifetime costs of screening tests in each strategy while outcomes relate to QALY gained. Incremental costs and QALY are reported to demonstrate the impact of parameter uncertainty on the costs and QALY outcomes for each analysis. Incremental costs and incremental effectiveness can be represented visually using a cost-effectiveness plane (CE plane) , which is used to present the probability of cases detected (Fenwick et al., 2001) (Fenwick and Byford, 2005) (Fenwick et al., 2006). The joint distribution of the costs and effectiveness from the Monte Carlo simulations (1000 iterations) are plotted on a cost-effectiveness plane in order to show the impact of uncertainty in the model parameters on the model outcomes in both expected incremental costs and effects. Uncertainty in the incremental outcomes is demonstrated when the results cross over the y-axis, representing both QALY gains (in the eastern quadrants) and QALY losses (in the western quadrants). Similarly, a spread through the origin passing through the x-axis represents uncertainty in the incremental costs of the intervention. The size of the spread also shows the extent of uncertainty in the costs.

Considering the deterministic model in the previous section, costs and effectiveness in both short and long term complications are seen to be slightly different in each strategy. When those results are plotted on the CE plane the distributions for each strategy overlap each other to such an extent that they cannot be distinguished clearly. Therefore, instead of presenting a distribution of cost and effectiveness on the CE plane, this model presents confidence intervals for each strategy, calculated using the percentile method (2.5 and 97.5 percentiles).

9.3.1 Short term complications in mothers and offspring

The costs and outcomes for each strategy, including their means and 95% CI's are presented in Table 9.6. The means of costs and outcomes from the Monte Carlo Simulation (1,000 iterations) for each strategy give slightly different results from the deterministic model. For example, S5 costs in the deterministic model are £1,202 and in the probabilistic model they are £1,180 (95% CI £918 – £1,463). Similarly, the means of the outcomes for S5 in the deterministic and probabilistic models are 1.9158 and 1.9158 (95% CI 1.8858 – 1.9404), respectively.

Table 9.6 The confident intervals for costs and QALY in short term complications

Strategy	Costs (£)			Effectiveness		
	Mean	Lower 95% CI	Upper 95% CI	Mean	Lower 95% CI	Upper 95% CI
S5	1,180	918	1,463	1.9158	1.8858	1.9404
S1 (NDS)	1,186	923	1,469	1.9161	1.8861	1.9406
S2 (NDS)	1,193	931	1,477	1.9160	1.8859	1.9405
S4 (NDS)	1,196	933	1,480	1.9160	1.8860	1.9405
S3	1,197	935	1,479	1.9162	1.8862	1.9407
S1 (PDS)	1,207	941	1,487	1.9159	1.8860	1.9404
S4 (PDS)	1,219	957	1,502	1.9155	1.8854	1.9401
S2 (PDS)	1,228	969	1,509	1.9152	1.8850	1.9396

9.3.2 Long term complications in mothers and offspring

Table 9.7 shows probabilistic results for costs and outcomes in terms of their means and 95% CI's for long term complications. The mean costs for each strategy in the table show similarity to the deterministic model. For example, the costs of S4 (PDS) range from £3,064 to £5,424 with a mean cost of £4,053. Likewise, there is uncertainty in the costs for long term complications when compared to the deterministic model (£4,088). On the other hand, the outcomes in both the deterministic and probabilistic models are slightly similar; 80.9736 and 80.9084 (95% CI 78.8925 – 82.7718), respectively.

Table 9.7 The confident intervals for costs and QALY in long term complications

Strategy	Costs (£)			Effectiveness		
	Mean	Lower 95% CI	Upper 95% CI	Mean	Lower 95% CI	Upper 95% CI
S4 (PDS)	4,053	3,064	5,424	80.9084	78.8925	82.7718
S3	4,213	3,300	5,682	80.9077	78.8954	82.7653
S1 (PDS)	4,236	3,341	5,692	80.9075	78.8952	82.7651
S2 (PDS)	4,308	3,351	5,692	80.9068	78.8913	82.7659
S1 (NDS)	4,358	3,416	5,877	80.8910	78.8806	82.7488
S4 (NDS)	4,457	3,459	5,868	80.8829	78.8748	82.7447
S2 (NDS)	4,497	3,489	5,915	80.8783	78.8731	82.7266
S5	4,622	3,528	6,388	80.8620	78.8511	82.7262

9.4 Decision uncertainty

The aim of calculating a Cost Effectiveness Acceptability Curve (CEAC) is to summarise and illustrate graphically the probability that a particular intervention or screening test is the optimal choice over a wide range of values for the ceiling ratio (λ). The CEAC is an appropriate way to present uncertainty in a cost-effectiveness analysis because it directly addresses uncertainty of the outcomes and costs of health care intervention (Fenwick et al., 2001). In multiple intervention decision, the CEAC for each screening test strategy can be established by calculating the proportion of iterations (from a Monte Carlo simulation) where the screening is optimal for a variety of willingness to pay values. The multi curve in CEAC of screening tests for GDM shows the decision uncertainty surrounding the cost-effectiveness of each strategy as discussed below.

9.4.1Short term complications in mothers and offspring

For short term complications, the CEAC show the decision uncertainty surrounding the cost-effectiveness of each strategy by plotting the probability of screening test strategies being cost effective against a range of ceiling ratios. For example, in Figure 9.3, at the ceiling ratio of £30,000/QALY the probability that S1 (NDS) is cost effective is 64%. Following this, for the remaining strategies S5, S2 (NDS), S4 (NDS), S3 , S1 (PDS), S4 (PDS) and S2 (PDS) the probabilities for their cost effectiveness are shown to be 31.3%, 3.9%, 0.3%, 0.5%, 0%, 0% and 0.01%, respectively. At the ceiling ratio of £30,000 per QALY, S1 (NDS) has the highest expected net benefit, but the error probability is about 36% (0.36 chances will be wrong decision). If the decision maker's ceiling ratio is below £20,000/QALY then S5 would be the screening test of choice since it is the cheapest. However, at the willingness to pay from £20,000/ QALY to £80,000, S1 (NDS) is cost-effective.

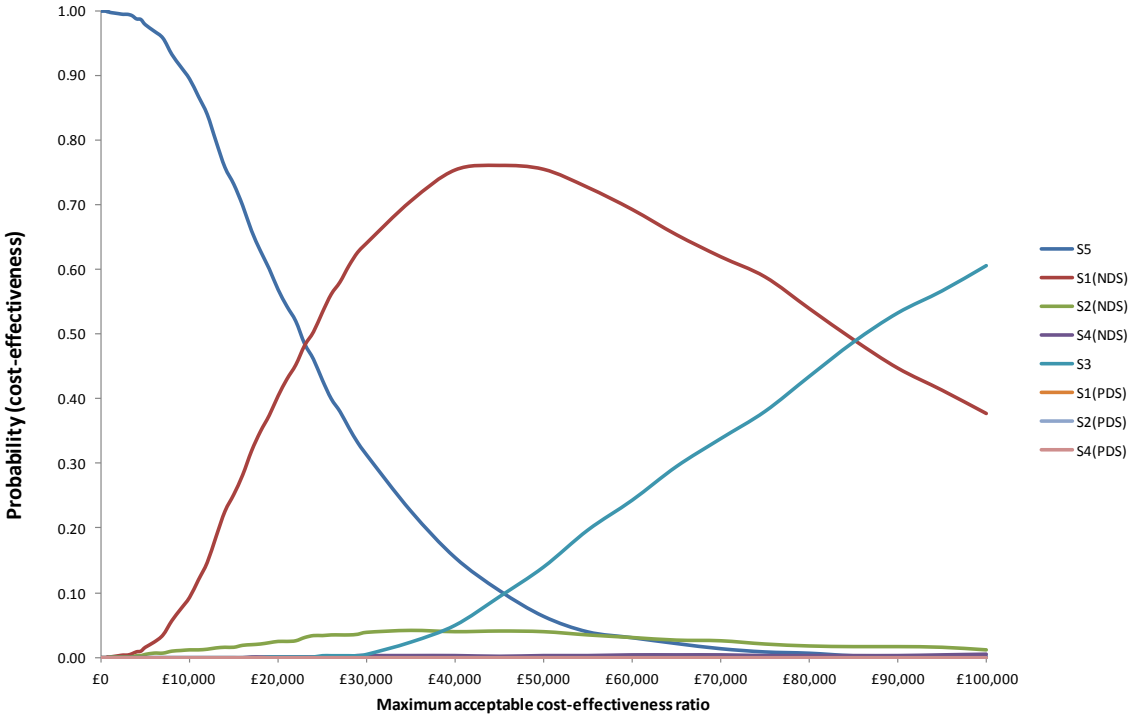


Figure 9.3 Cost-effectiveness acceptability curves for short term complications

9.4.2 Long term complications in mothers and offspring

The uncertainty in cost-effectiveness estimated for long term complications is illustrated in a CEAC in Figure 9.4. At a threshold of £30,000/QALY, the CEA illustrates that the probability that S4 (PDS) will be cost-effective is approximately 55.8%. At the same threshold, the probability that S3, S1 (NDS), S5, S2 (NDS), S4 (NDS), S2 (PDS), S1 (PDS) will be cost-effective are approximately 37.6%, 0.3%, 1.4%, 0.6%, 0.1%, 0.4% and 0.0%, respectively. However, the probability that S4 (PDS) will be cost-effective in the CEAC remains at approximately 55.8%.

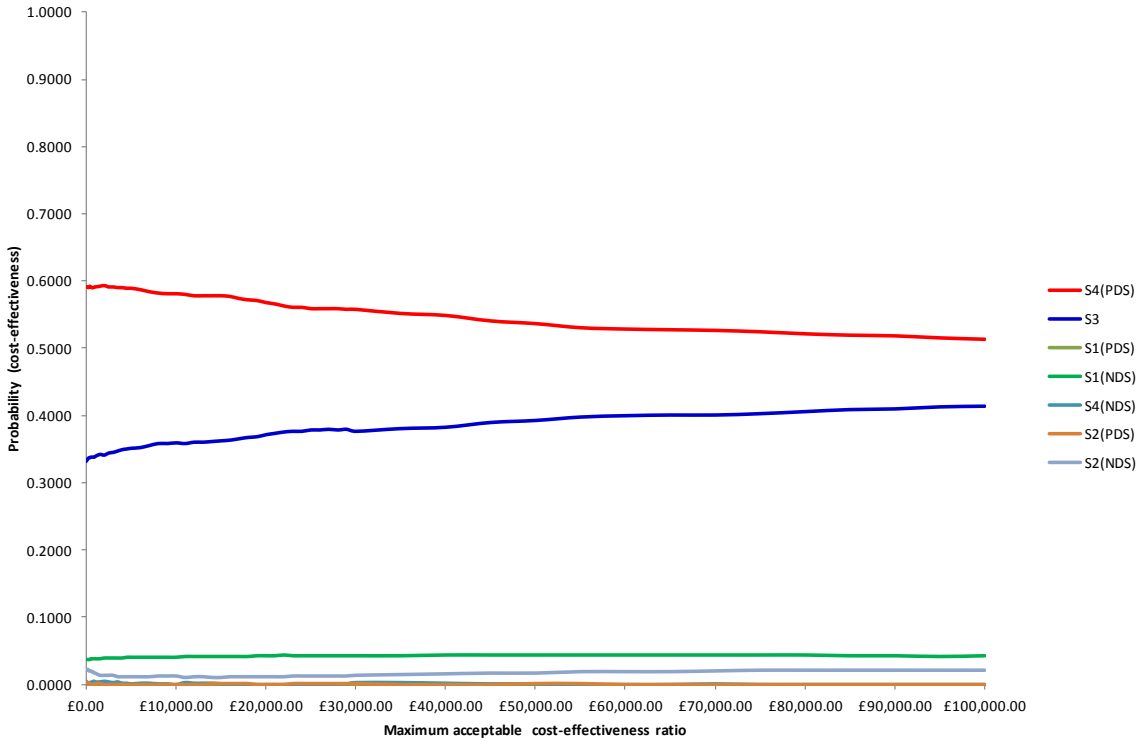


Figure 9.4 Cost-effectiveness acceptability curves for long term complications

9.5 Value in further research: EVPI and EVPPI

The probabilistic sensitivity analysis results have been presented in previous sections and the decision uncertainties for each of the analyses have also been explored, however to demonstrate the probability that each intervention is cost-effective at different thresholds it is necessary to consider two issues, namely, knowing the decision uncertainty and current evidence, should the screening test be adopted or not and is additional research essential in order to support the decision (Briggs et al., 2006). If such questions are left unanswered, it may be the case that decision makers attempt to interpret the cost effectiveness results in terms of how to make the decision to adopt or reject the screening tests given the uncertainty. The outcomes from the probabilistic sensitivity analysis in the previous section can help inform funders and decision makers to answer those questions about how to interpret the results for each screening test and the appropriate next steps to take.

The potential value of undertaking further research is estimated using a Bayesian Value of Information Analysis (VOI) and is discussed in this section. Firstly, the expected value of perfect information (EVPI) is estimated, the value of eliminating all the uncertainties within the model, providing a maximum value for the return on further research. The EVPI was calculated using the probability of cost-effectiveness for each strategy, and were generated in the CEAC calculation within a range of values at

intervals of £500 from zero to £100,000/QALY. It is important to represent what this EVPI per patient is in terms of the relevant patient populations who would benefit for the additional information. Representing the EVPI per decision in terms of the patient population, the population value for EVPI, gives an idea of the upper limit for expenditure on future research into the decision question. It is calculated by multiplying the patient population (both present and future) that is able to benefit from the information, by the difference between a decisions expected value (the greatest net benefit expected) with current information and with perfect information (Briggs et al., 2006).

In order to reduce uncertainty in the cost effectiveness decision, it is necessary to consider what parameters are driving uncertainties in the model. In this respect, the expected value of perfect parameter information (EVPPI) is used to distinguish parameters for which it would be valuable to have more accurate estimates. Parameter uncertainty for one or a group of parameters can be eliminated instead of the EVPI for all parameters being estimated at once. The EVPPI is calculated from the difference between the maximum net benefit calculated from current estimates that involve some uncertainty and the expected net benefit that is calculated from partial perfect information. This provides the proportion of general uncertainty furnished by any single or group of parameters and gives a guide to where research should be focused for greatest efficiency.

9.5.1 EVPI in short term complications

The EVPI results showed that at a willingness to pay threshold of £30,000/QALY, the value of perfect information for short term complications is £1.06. This EVPI is very low; however an explanation for this may be due to a lack of decision uncertainty as seen in Figure 9.3 in the previous section. At a threshold of £30,000 per QALY, the probability that S1 (NDS) would be cost-effective was only 64%. The EVPI equates to the opportunity loss form choosing the optimal strategy in a situation where the optimal strategy would have been wrong had perfect information been provided. This therefore means that there is approximately a 36% probability that S1 (NDS) is the wrong decision. Thus it can be inferred that the opportunity loss of choosing the screening strategy is high at a willingness threshold of £30,000/QALY.

Having estimated EVPI per patient, the population EVPI for one year is calculated using population estimates. In order to determine the population value for EVPI, the patient population over the lifetime of the technology must be taken into account in terms of the relevant patient population who would benefit from the screening

tests. The annual pregnancy population is estimated to be 60,000 mothers (NHS, 2013). With regard to screening tests for GDM, this study assumes an effective technology life of 10 years with new patients eligible for treatment each year. Therefore, a timeframe of ten years was applied and discounted at 3.5% (NICE, 2008c). The EVPI for the population has the same relationship to the CEAC calculation (Figure 9.3). Figure 9.5 details the results from the expected value of perfect information analysis (EVPI) at a population level for short term complications. At a willingness to pay threshold of £30,000/QALY the expected value of perfect information for the populations is approximately £494,990. If a monetary threshold of £100,000 per QALY were applied, then the population value of further research would be £658,717. At the willingness to pay £20,000/ QALY, the line of population EVPI starts to drop and then rises again at a willingness to pay of £80,000. This drop relates to the CEAC for short term complications in Figure 9.3, in which the S1 (NDS) strategy shows the highest probability of cost-effectiveness between the willingness to pay values stated above. For example, at a willingness to pay of £20,000/QALY, the probability that S1 (NDS) is cost effective is 64% and the probability that the other strategies are cost effective is 36%. Therefore, at a willingness to pay of £20,000/QALY a peak in population EVPI can be seen in the graph in Figure 9.5 it shown uncertainty at this willingness to pay. However, there is still considerable uncertainty surrounding this decision, and at a willingness to pay value of £30,000/QALY the probability that S1 (NDS) is cost-effective is approximately 64%.

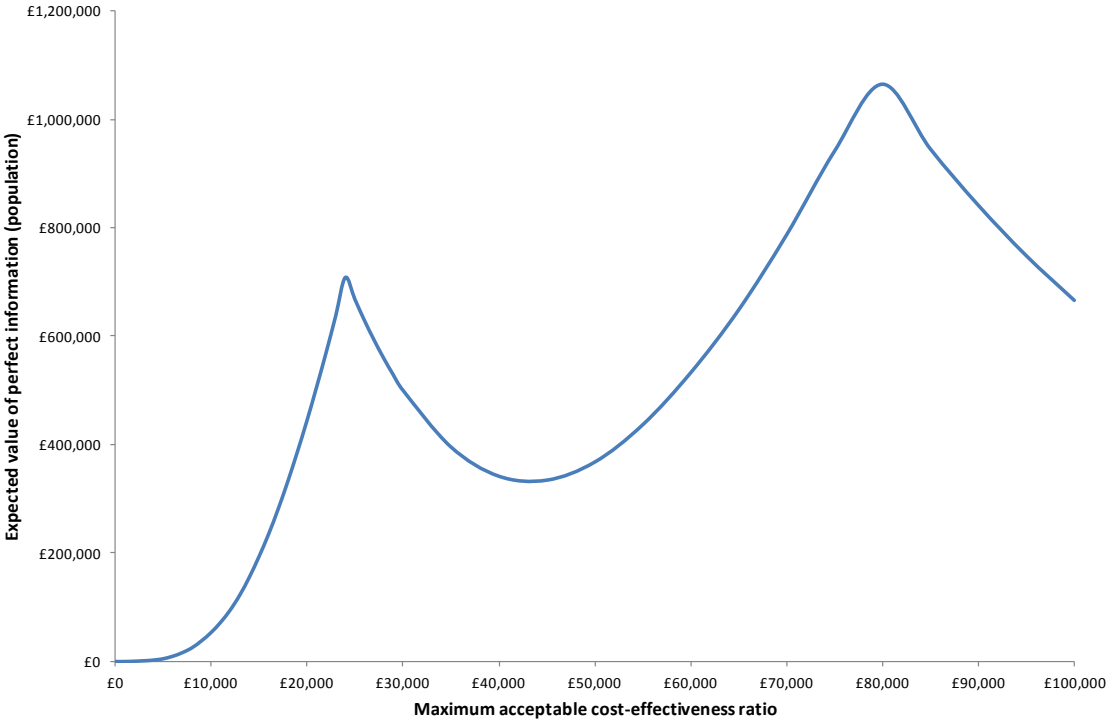


Figure 9.5 EVPI for short term complications – population level

At a willingness to pay threshold of £30,000/QALY, the population EVPI was approximately £500,000. Therefore it is potentially worthwhile undertaking further research to explore screening tests for GDM if further research costs less than £50,000. If it costs more than £50,000, further research is not worthwhile.

9.5.2 EVPI in long term complications

The EVPI results indicated that at a threshold of £30,000/QALY, the expected value of perfect information per patient is £229.14. This value can be explained by taking into consideration the decision uncertainty shown by the CEAC in Figure 9.4. The CEAC for long term complications showed that at a threshold of £30,000/QALY, the S4 (NDS) would be approximately 55% cost-effective, with a 45% probability that a wrong decision would be made.

In order to determine the overall population value of EVPI, the analysis for long term complications was performed based on the same technology time frame and population of pregnant women in Scotland (60,000) (NHS, 2013). Figure 9.6 illustrates the results from the expected values of perfect information at the population level for long term complications. When scaled up to reflected this value to the population, the EVPI results showed that at a monetary threshold of £30,000/QALY the expected value of perfect information per populations for GDM screening tests in mothers is

approximately £107 million. This means that further research is worth £107 million as shown in Figure 9.6. The EVPI for the population has the same relationship to the CEAC calculation in Figure 9.4. In the CEAC, the probability that S4 (PDS) will be cost-effective remains at approximately 55% from willingness to pay threshold of £0/QALY to £100,000/QALY. Therefore, the line of population EVPI starts to rises up as a straight line at a willingness to pay threshold of £0/QALY to £100,000/QALY.

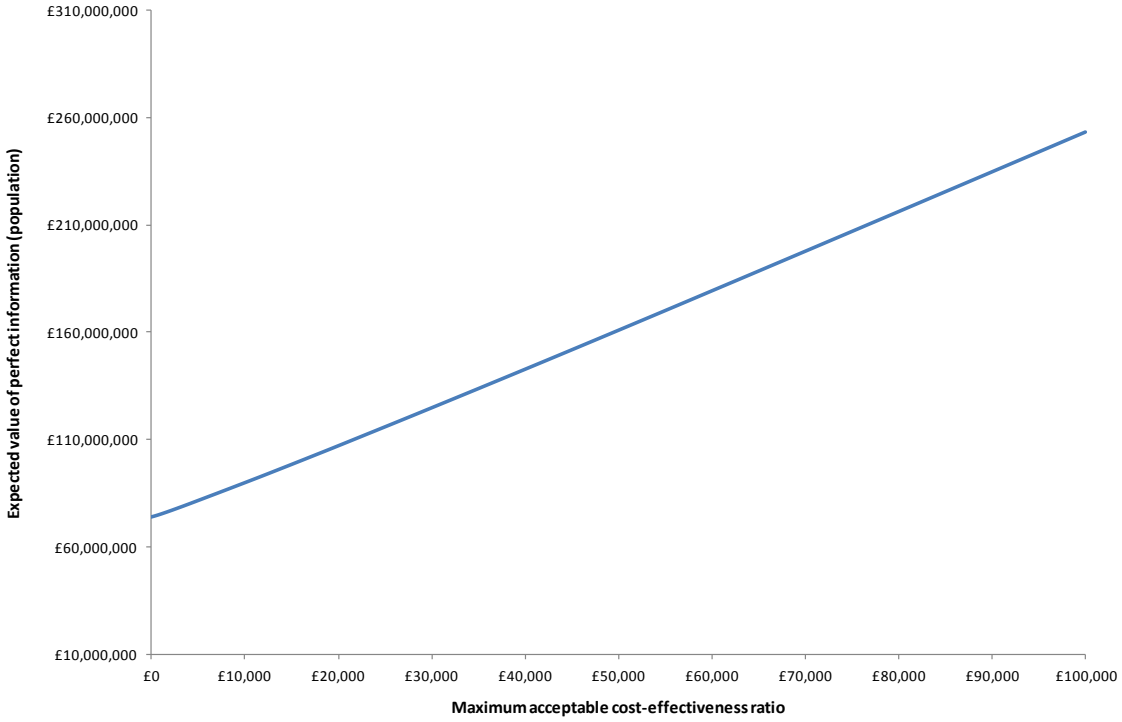


Figure 9.6 EVPI for long term complications - population level

The EVPI results show that it is worthwhile collecting more information about the use of S4 (PDS), as a selective screening test strategy for GDM. At a willingness to pay threshold of £30,000/QALY the population EVPPI was £107 million. Even at a willingness to pay of zero, the population EVPPI is still extremely high at £75 million. With such a high population EVPI, it can be concluded there is potential in undertaking future research to reduce the uncertainty that surrounds the question of screening tests for GDM using S4 (selective screening).

9.5.3 EVPPI in short and long term complications

The EVPI per patient of £1.06 translated to population level EVPI £494,990, at a willingness to pay threshold of £30,000/QALY in short term complications. Whereas, in long term complications, the EVPI per patient of £229.14 translated to population level EVPI £107 million, at a willingness to pay threshold of £30,000. Considering the EVPI at

the population level, there would be value in investigating further information on screening tests for GDM. In terms of which parameters would add the most value in this additional information and how it should be collected, the expected value of perfect parameter information (EVPPI) is used; understanding what drives the uncertainty is necessary to establish the type and perhaps the scale of future research. The EVPPI analysis reports results in terms of the value per patient and these are presented in terms of the pertinent patient population that may benefit from this supplementary information. The parameters are arranged in two main groups based on short and long term complications, as summarised in Table 9.8. Each group consists of sub-groups that contain individual parameters specific to screening tests for GDM. For short term complications, the five sub-groups indentified were prevalence, diagnostic test accuracy, costs of treatment, clinical effectiveness and utility. Whereas, the sub-groups for long term complications include prevalence, cost for treatment, life table, and utility. Individual parameters and groups of parameters were selected in order to find out which parameters may be causing the high EVPI.

Table 9.8 EVPPI parameters

Parameter Groups	Parameters
Short term complications	
Prevalence	GDM
	GDM with and without risk factors
Diagnostic test accuracy	Random plasma glucose: sensitivity & specificity
	Fasting plasma glucose: sensitivity & specificity
Cost of treatment	Normal delivery
	Short term complications in mothers
	Short term complications in offspring
Clinical effectiveness	GDM with treatment (TP)
	GDM without treatment (FN)
	No GDM with treatment (FP)
	No GDM without treatment (TN)
Utility	Normal delivery
	Short term complications in mothers
	Short term complications in offspring
Long term complications	
Prevalence	Development of type 2 DM in mother with and without GDM
	Development of type 2 DM in offspring with mother with and without GDM
Cost of treatment	Type 2 DM with no complications : in and out patient costs
	Type 2 DM with no complications : in and out patient costs
Life table	Death rates
	Death rates with DM
Utility	UK Population Norms for EQ-5D
	HRQoL in type 2 DM
	HRQoL in type 2 DM with complication

The population EVPPI was estimated based on the population of pregnant women, 60,000 (NHS, 2013). The EVPPI analyses, in terms of the values per decision, are presented based on the relevant patient population that benefits from the additional

information. The EVPPI was run using 100 X 100 iterations for each of the individual parameters and groups of parameters, at a willingness to pay threshold of £30,000/QALY. For the individual parameters and sub-group parameters for both short and long term complications, the EVPPI analyses show that there is zero value in undertaking further research. It is clear that these parameters may have some uncertainty surrounding them, but not enough to warrant any further information. It is completely reasonable for EVPPI for all individual parameters to be zero, but as a group they may be sizeable (Briggs et al., 2006). The sum of individual parameter groups does not necessarily equal the EVPPI of all groups together. The reason for this is that when the individual parameter groups are considered separately they may not resolve in a way that has enough impact on the difference in net-benefit for the decision to be changed. Parameters for both short and long term complications have zero individual EVPPI, but when combined they may have the effect of reducing uncertainty. With EVPPI of zero for individual parameters, it is not worthwhile to collect information on individual parameters. However, for example, if the prevalence of the GDM group had an EVPPI per patient of £5, then this translates to population values of approximately £50,000. Consequently, it may be possible to undertake research on this parameter at a cost lower than £50,000. However, all individual parameters have zero results and so it is clear that the parameters may have some uncertainty surrounding them, although the uncertainty is extremely low and not enough to warrant any further information.

The EVPPI was recalculated based on the two groups of parameters; short and long term complications. For those two groups of parameters, the EVPPI analysis reported a value of zero for short term complications and for long term complications, the value of EVPPI was worth £1.67 per patient. Short term complications reported a value of zero because the low EVPI for short time complications was £494,990, with willingness to pay a threshold of £30,000/QALY. The EVPI for population level in short term complications was not a high value when compared with the EVPI for population level in long term complications (£107 million). Another reason for a zero value is that the individual parameter groups are different for short and long term complications. The EVPPI groups for both short and long term complications result in different values and the uncertainty in the individual parameters. For example, the individual parameter groups for short term complications included diagnostic test accuracy and clinical effectiveness; whereas the individual parameter groups for long term complications did not include these groups. Therefore, EVPPI for all groups in both short and long term are not directly comparable as they are composed from different components. If this is scaled up to the population level it gives a population EVPPI of £784,042 for long term complications, as shown in Figure 9.7. These results shows that there is greater

uncertainty with respect to long term complications and that collecting information on long term complications is likely to be worthwhile.

Clearly it is the parameters for long term complications that are driving uncertainty in the model, and leading to uncertainty in the cost-effectiveness decision. Moreover, it can be seen that it would be best for EVPPI analysis to first be conducted on individual parameters or small groups of parameters, with the grouping of parameters being of particular importance if they are correlated. Also even though it may be quite possible for EVPPI to equal zero on all the individual parameters, as groups they may have substantial value.

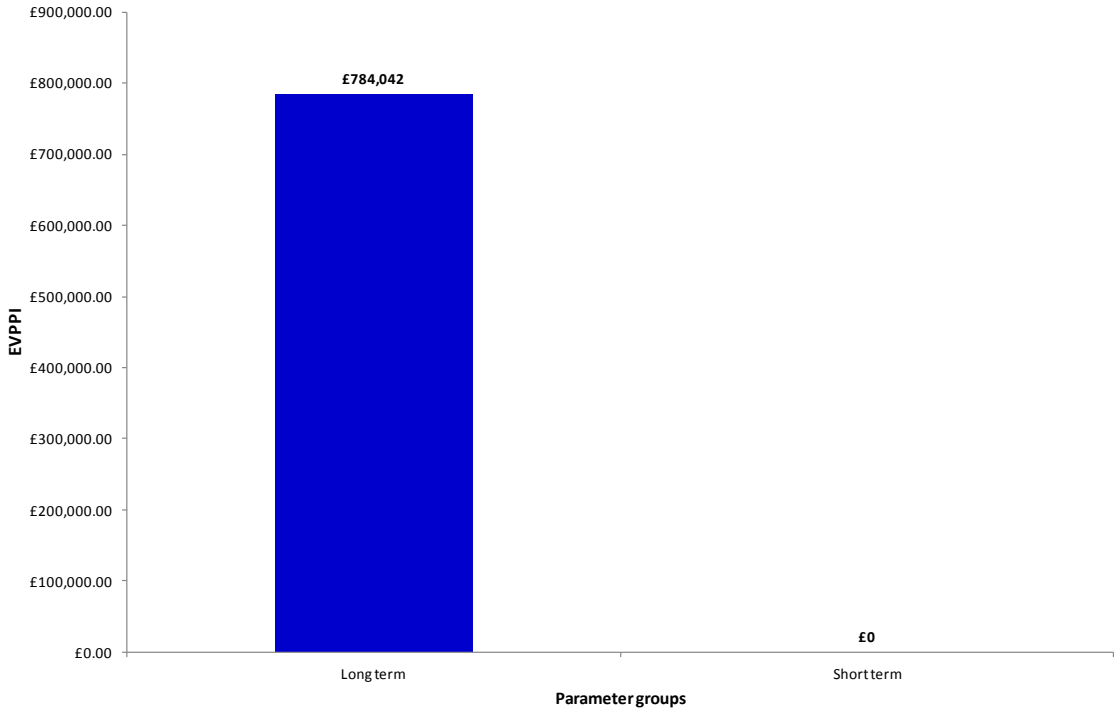


Figure 9.7 EVPPI for short and long term complications – population level (£30,000)

This model is the first economic evaluation of screening tests for GDM that has applied VOI analysis in order to decide whether or not to adopt screening tests and whether or not to carry out further research. There is uncertainty about the probability of the parameters used in long term complications, such as costs, prevalence of developing type 2 DM, and QALY. This means that the decision is uncertain and that there is a probability of decision error. The EVPPI showed that, at the willingness to pay threshold of £30,000/QALY, it is worthwhile reducing uncertainty in all parameters for long term complications. Given the result of the EVPPI, the type of further research that would be required is likely to be a trial to evaluate the development of type 2 DM in mothers with GDM and in offspring born to mothers with GDM. A randomised controlled trial (RCT)

would provide strong evidence as to whether long term complications in mothers with and without GDM develop type 2 DM. However, any such further research on long term complications would require a lengthy study period and incur high costs.

9.6 Discussion

Results for decision models for which probabilistic sensitivity analysis has been applied can be estimated by VOI analysis. This chapter explored uncertainty in the results of screening tests of GDM and demonstrated the use of the VOI method. The results can be examined in terms of decision uncertainty to give meaningful recommendations for further research. The results for each of the analyses are discussed and summarised.

9.6.1 Short term complications

The cost-effectiveness planes for short term complications showed that there was a lot of uncertainty surrounding all strategies. Of all the screening tests for GDM, S1 (NDS) is the most cost-effective, and at a willingness to pay of £30,000/QALY was 64% cost-effective, representing a large uncertainty surrounding the decision. In considering the value of further research, the EVPI value was very small, reflecting the low decision uncertainty. In addition, individual and sub-group parameters for short term complications had zero values in the EVPPI analysis. Therefore, given current evidence it is certain that the use of S1 (NDS) as recommended by SIGN 2001 is cost-effective, and that there is no value associated with the collection of further information. It is surprising that short term complications in screening tests for GDM are not driving uncertainty in the model. Individual parameters such as prevalence rate and diagnostic test accuracy should drive uncertainty in short term complications. However, the EVPI shows that perfect information would not change the cost-effectiveness decision, and therefore there is no value in conduction further research.

In short term complications, S1 (NDS) is a universal screening strategy with a two step approach that screens by random plasma glucose (RPG) followed by a 75g oral glucose tolerance test (OGTT) when the screening test is positive (SIGN, 2001). For populations with low prevalence of GDM at 3.5%, this model found that universal screening is a cost-effective strategy for short term complications. S1 (NDS) shows the lowest screening costs at £7 with an expected case identification rate of 2.2%. In contrast all of the PDS strategies were able to classify 3.5% of mothers as having GDM, which is close to the prevalence rate used by this model. Therefore, the screening with the highest case identification rate is not necessarily the most cost-effective in short

term complications. There may be some desirable trade-off between detection, unnecessary testing and treatment for GDM.

In short term complications, when comparing PDS and NDS, NDS is more cost-effective than PDS as illustrated in Figure 9.1. In the cost-effectiveness plane for short term complication, all NDS screening test strategies were presented at the bottom of the graph. The NDS results are due to an error in FN tests, in which all patients with disease are undiagnosed and do not receive any treatment to prevent short term complications. On the other hand, the PDS strategies presented with a zero probability for FN. In FN, mothers with GDM do not receive any treatment, however the treatment of GDM reduces serious adverse complications in both mothers and offspring (Crowther et al., 2005), and so mothers who show FN results cost more money due a greater need to treat for both short and long term complications, in comparison to the other results; TP, TN and FP. Therefore screening test strategies with low screening test costs and a low number of FN results are cost-effective.

The cost-effectiveness estimates in this model are similar to the estimates in a number of economic evaluations previously mentioned concerning the screening of GDM and short term complications. The results in the present model conclude that the sequential screening test or two step approach in which patients that test positive in the first test (screening test) undergo a second test (diagnostic test), was the most cost-effective strategy. Several economic evaluations of screening tests for GDM reported sequential screening tests as being cost-effective (Poncet et al., 2002) (NICE, 2008b) (Round et al., 2011). A study in cost-utility for different screening tests in England and Wales reported that a sequential strategy of FPG followed by OGTT is the optimal strategy for short term complications, with 100% sensitivity and specificity (Round et al., 2011).

The RPG test has reasonable sensitivity with low levels of prevalence in the population. This means that the additional number of GDM cases that would be missed is small. It also represents a lower cost strategy than testing all women with 75g OGTT. Nevertheless, confirming an RPG positive with 75g OGTT is still cost-effective as the cost of the test is offset completely by the savings made from not treating FP cases. If it is assumed that 75g OGTT diagnostic test acceptance is not 100%, the number of cases missed will be established by combining test sensitivity and acceptability. From this, it is clear that when the assumption of 100% sensitivity and specificity is relaxed, OGTT itself is not therefore used as a cost-effective screening test. Regarding diagnostic test accuracy in this study, fasting plasma glucose (FPG) can classify more

GDM cases than RPG, however when there is low risk of GDM in the population, the probability of case identification by FPG and the cheaper RPG is small.

This clarifies why RPG followed by 75g OGTT is cost-effective. With 100% sensitivity and specificity, as part of a combined testing strategy where the woman has already had one positive result with RPG, it is cost-effective to use 75g OGTT to confirm a positive RPG or GCT before treatment. This is because the use of the diagnostic test results in savings made from not treating FP. The probability of FP cases in NDS is zero; however the probability of FP in PDS for each of the strategies does exist. Therefore, Universal screening with RPG and 75g OGTT for mothers and offspring represents the best screening test strategy when considering costs and outcomes over one year. In short term complications, the time horizon may not reflect the clinical outcome in both mothers and offspring with GDM.

9.6.2 Long term complications

For long term adverse complications, the CEAC indicated that at a willingness to pay threshold of £30,000/QALY the probability of S4 (PDS) being cost-effective is 55%. The probability that S4 (PDS) is cost-effective in the CEAC remains the same at willingness to pay thresholds of between £30,000/QALY and £100,000/QALY. There is a lot of uncertainty surrounding all of the strategies. At a population level the EVPI is £107 million for long term complications and £494,990 for short term complications, and therefore there is potential worth in collecting further information to inform the decision regarding whether to use S4 (PDS) in the future. The population EVPI analysis of long term complications demonstrates that further research is worthwhile. As a research recommendation, outcomes of the EVPI are useful in terms of being able to set upper limits on future research costs, such that research has to have a lower cost than the EVPI for it to be deemed of potential worth.

The EVPPI outcomes for various individual parameters and sub-groups of parameters equal zero with the exception of the group of long term complications which was £1.67 per patient, and £784,042 at the population level. To help make decisions as to what type and scale of research is most suitable, EVPPI analysis can be employed. It takes into consideration which parameters drive uncertainty in a cost-effectiveness decision, such that by determining the relevant parameters, the most suitable type of research can be identified. Long term complications are driving the uncertainty in this model. Further research should focus on whether long term complications in mothers and offspring with and without GDM develop type 2 DM. However, such further research would require a lengthy study period and may incur high costs. However, large scale

RCT would not necessarily be required; indeed the type of research would be dependent on which parameters required additional information. Performing EVPI and EVPPI with the use of VOI analysis, there is potential worth in making most use of outcomes and in making informed recommendations to those providing funds as well as decision makers.

For short term complications, universal screening with NDS is a cost-effective strategy, whereas for short and long term complications, PDS with selective screening tests is cost-effective. For short term complications, a diagnostic test reduces the costs involved in treating FP, as there is a zero probability of FP in NDS. As there are no costs incurred in the treatment of FP, NDS is the dominant strategy for short term complications. Moreover, the high cost and low utility come from the probability of FN in NDS itself. FN is considered to be worse than FP, as mothers with GDM will not get any treatment and because the treatment of GDM reduces serious adverse complications in both mothers and offspring (Crowther et al., 2005). Mothers who show FP results will cost more money due to unnecessary treatment of short term complications.

When comparing PDS and NDS, PDS is more cost-effective than NDS. An explanation follows as to why PDS replaces NDS and selective screening replaces universal screening as the most cost effective strategy when considering long term complications as opposed to short term complications. To understand the diagnosis outcomes more clearly from both NDS and PDS, it is advisable to look at the probabilities of each test's accuracy in section 7.6.1. The main reason that PDS is dominant for GDM screening tests is because it is able to detect all mothers with a true disease state due to the high sensitivity and specificity of the tests used in the model (equal to prevalence of the disease in populations, 3.5%), and so the probability of FN results for PDS equals zero. Moreover, the PDS strategies involve FP test result errors whereby all patients receive unnecessary treatment during gestation to prevent short term complications. It is therefore recommended that mothers in the FP group (over-diagnose) for short term complications receive postpartum screening for type 2DM, in the same way that TP patients do. In other words, all patients in the TP and FP groups are screened for type 2 DM which allows for the early detection of the disease and its treatment, preventing the onset of complications.

When both short and long term complications were included in the same model to analyse the cost-effectiveness of screening tests for GDM, costs and outcomes were presented in terms of lifetime QALY in type 2 DM. According to this model, S4 (PDS) that involved screening for GDM with risk factors at initial visit followed by 75g OGTT was found to be the most effective. For long term complications, mothers with GDM are

at an increased risk of developing type 2 DM (Bellamy et al., 2009). This model assumes all mothers with GDM receive postpartum screening for type 2 DM at three intervals over 10 years. With this assumption, all postpartum mothers with GDM are screened for type 2 DM. Therefore, mothers with GDM receive a chance to be screened for type 2 DM. When the results of screening tests are positive those mothers receive the appropriate treatment to prevent type 2 DM complications. Diabetes without treatment results in high glucose levels in the blood, which can damage nerves, organs and lead to serious illness. Diabetes with complications leads to poor health, premature mortality, and to a reduction of life expectancy and quality adjusted life expectancy. Therefore, screening tests for GDM that can capture all true disease in a population may reduce the costs of treatment of type 2 DM complications and improve QALY in mothers with GDM who develop type 2 DM. Thus, a screening strategy with PDS that can detect all patients in the population and does not show any FN results is cost-effective. Screening with the highest case identification rate is the most cost-effective strategy for long term complications. Therefore, with a GDM prevalence of 3.5%, it is cost effective for all pregnant women that screen positive for risk factors at first visit to receive treatment care for GDM, and for women with negative risk factors to undergo diagnostic tests with 75g OGTT.

The cost-effectiveness estimates in this model are similar to the estimates in a number of economic evaluations previously mentioned concerning the screening of GDM and long term complications. This model demonstrates that the selective screening approach to GDM screening and diagnosis is cost-effective compared to a no-screening strategy as long as GDM diagnosis provides an opportunity for early intensive intervention and prevention of overt DM (Poncet et al., 2002) (Nicholson et al., 2005). Selective screening with risk factors followed by 75g OGTT was cost-effective for long term complications. As discussed previously for short term complications, it is cost-effective to use 75g OGTT to confirm positive results for either risk factors, FPG or RPG rather than using 75g OGTT alone. Similarly, two studies of the cost-effectiveness of screening tests for GDM stated that the most favourable ICER was the screening of high-risk pregnancies with test techniques such as 50g OGTT (Poncet et al., 2002) and 100g OGTT (Nicholson et al., 2005). Moreover, a further two studies in the UK suggested that the optimal strategy varied according to the women's risk factors (Round et al., 2011) (NICE, 2008b). However, in the low risk factors group (GDM risk is < 1%), the most effective strategy is to do nothing (Round et al., 2011). Therefore, screening based on individual risk is potentially more sophisticated and cost-effective to that based on an approach to risk factor screening that is widely discussed in the literature. In addition, this model was designed to include a no-screening strategy specifically so that it could be determined under what circumstances screening is cost-effective. The

results of the present model concluded that sequential screening tests or the two step approach, in which patients that test positive in the first test (screening test) undergo a second test (diagnostic test), also was the most cost-effective strategy as discussed in Section 9.6.1.

To this knowledge, this model assesses the costs and effects associated with GDM over the course of a lifetime, which is similar to other economic evaluations of screening tests for GDM (Lohse et al., 2011) (Werner et al., 2012) (Marseille et al., 2013). Several other studies have examined the cost-effectiveness of GDM treatment, but only in short term complications (Poncet et al., 2002) (Nicholson et al., 2005) (NICE, 2008b) (Round et al., 2011) (Mission et al., 2012). This model however, confirms and high-lights the need for long term intervention for postpartum screening, to curb progression to diabetes if any GDM screening program is to be cost-effective. In addition, the strength of this model is that it includes the quality of life of the infants after deliver in relations to short term complications and their life expectancy as they develop type 2 DM through childhood and young adulthood. This economic evaluation model included screening tests for GDM, management during pregnancy and postpartum screening in order to assess the cost-effectiveness of screening tests for GDM.

9.7 Conclusion

This chapter has presented a decision analytical model to estimate the cost-effectiveness of screening tests for GDM. To calculate the cost-effectiveness of diagnostic testing, modelling procedures are required for which intermediate measures associated with assessments of diagnostic tests are translated into short and long term outcomes and costs. This model used a decision tree to model the short term complications over a 1 year period and long term complications over a life cycle. The model was populated by using a combination of data from published literature and expert opinion to yield parameter uncertainty. Therefore, applying a decision analytic model and a VOI technique allows decision makers to make informed decisions as to whether a new intervention should be adopted or rejected based on current evidence.

Chapter 10 Main findings, Policy implications, limitations and future research

10.1 Introduction

The overall aim of this thesis is to study the cost-effectiveness of screening and diagnostic tests for GDM in Scotland. The narrative and systematic review, based on economic evaluations, diagnostic test evaluations and screening tests of GDM, are addressed at the beginning of this thesis. The systematic review of economic evaluations was performed to undertake a structural review and critical appraisal of the methods used in the cost-effectiveness analysis of screening tests for GDM and to inform the development of the model. Based on the critical assessment of the literature in the first six chapters, the empirical analyses undertaken in chapter 7, chapter 8 and chapter 9 presented issues in the research field that have not been dealt with previously in the literature, using a combination of test results for both the negative dominant strategy (NDS) and the positive dominant strategy (PDS). This thesis has, for the first time, evaluated the economic evaluation of NDS, PDS and the value of information (VOI) for screening tests for GDM.

Section 10.2 of this final chapter summarises the main findings of this thesis. Subsequently, in section 10.3, potential policy implications are outlined, followed by a discussion of possible limitations of the analyses in the thesis and generalisability in Section 10.4. Lastly, section 10.5 provides a presentation of how this work could be taken forward in the future.

10.2 Main findings

Models of screening tests for GDM were developed in order to estimate costs and outcomes in both mothers and offspring for four alternative strategies versus a strategy of “no screening”. In this economic evaluation, decision trees were designed to assess the five strategies of screening for GDM were constructed based on the combinations of test approaches for NDS and PDS. Moreover, the analyses were performed from a health care provider’s perspective (NHS).

10.2.1 The combination of test results

The thesis has shown that the most appropriate model makes use of a combination of tests as either “negative dominant strategy” (NDS) in which patients who tested positive after an initial screening test are brought back for a second test to further reduce false positives, or “positive dominant strategy” (PDS) whereby patients that test negative in the initial test are tested a second time to identify any missed true positives. As such the NDS and PDS approaches allow the clinician to consider test results in terms of the differences in false negative (FN) and false positive (FP) test results. Combining the tests in terms of NDS and PDS involves a trade-off between sensitivity and specificity. With a two-step approach, PDS benefits from sensitivity, and results in an overall larger number of positive outcomes, both true positive (TP) and FP, and a reduction in negative outcomes. This is at the expense of indentifying fewer TN and also gives a greater proportion of FP (trading-off specificity for improved sensitivity). The net effect of testing the negatives from the first test is finding more TP (net sensitivity will be higher than sensitivity from individual test). In terms of treatment for patients, the proportion of FP must be considered, as they can lead to over treatment. Whereas the NDS approach favours specificity and results in a larger proportion of true negative (TN) and FN. Therefore, NDS benefits from the specificity of both tests, at the expense of lower overall sensitivity (trading-off sensitivity for improved specificity). The net effect of testing the positive from the first test is finding more TN (net specificity will be higher than specificity from individual test). Again in terms of treatment for patients, the proportion of FN must be considered, as this is likely to lead to missed treatment. The trade-off between sensitivity and specificity is at the heart of screening and diagnostic tests and the accuracy of the tests, in terms of sensitivity and specificity, is an important consideration for clinicians.

10.2.2 The prevalence of GDM: Systematic review

Prevalence of GDM is one of the most important parameters in cost-effectiveness analyses. However, there remains considerable uncertainty regarding its estimate. The prevalence estimates for GDM in the systematic review reported that the pregnancies of 76,312 women ranged from 1.35% to 12.80%. However, considerable heterogeneity in the prevalence of GDM was identified and so pooled estimates were not calculated. Variation in GDM prevalence is likely explained by ethnicity and diagnostic screening strategies. This is because mothers that belong to a high risk ethnic group are more likely to test positive which in turn leads to higher prevalence rates compared to other populations. Indeed, screening ethnic groups that have a high risk of developing GDM

can result in high prevalence estimates, ranging from 8.5% - 12.8%. Other important factors are whether universal or selective screening protocols for GDM are adopted and whether 75g or 100g OGTT are used. The impact of these factors on prevalence estimates should be further investigated as they may be acting as proxies for other influences on prevalence. Furthermore, accurate decision models require accurate estimates of prevalence and decision model outputs are highly sensitive to the prevalence. Additionally, in economic evaluations of screening tests prevalence is required as a probability in the construction of decision trees.

10.2.3 Costs and cost-effectiveness analysis of screening tests for GDM: Systematic review

Costs associated with the screening and management of GDM vary widely by country, ranging approximately from £2.42 to £50.9 per case detection, and are dependent on the tests used, the screening approach, how the costs are calculated and the prevalence of GDM in the population. The difference within the reviewed studies in terms of comparators, interventions, outcome measures and the analysis of uncertainty, make meaningful comparison between the studies very difficult. The available information on cost effectiveness provides strong evidence in support of the use of risk factor screening tests for GDM in high-risk groups, such as family origin with a high prevalence of diabetes, previous gestational diabetes and obesity. However, the review of the CEA study has significant implications for future research and policy making. Two key findings followed from the review. The first was that CEA of screening tests for GDM should use multiple time horizons, such as one year and over a lifetime and as such case identifications and short term complications (one year time horizons) are not the endpoint outcomes for GDM screening tests. A major reason for introducing screening programmes in the population is that it is important to initiate the appropriate treatment after positive cases have been identified with respect to short term time horizons. However, the effect of screening and the treatment of GDM, that includes blood glucose monitoring, oral medication and regulating insulin, is a reduction in the likelihood of developing short term complications and also long term complications, namely type 2 DM in both mothers and offspring. Indeed, a potentially essential outcome for the detection of GDM is the identification of women who are at high risk of developing type 2 DM or at high risk of GDM in subsequent pregnancies. Long term consequences are therefore appropriate outcomes for the CEA of screening tests for GDM and the model for economic evaluation for GDM should account for the effectiveness of postpartum screening for type 2 DM over the long term. The second finding was that no studies in the review presented the results of value of information (VOI) for screening tests for

GDM. VOI can be used to make meaningful recommendations to decision makers, who can then make informed decisions as to whether a new intervention should be adopted or reject based on current evidence, or whether further information is required to help make the decision. Its use is important as a decision on the most suitable screening test must be based upon and supported by all available knowledge.

10.2.4 Case identification

The decision analytic model studied cost effectiveness by expressing consequences in terms of natural units as cases of GDM detected. Decision trees for case identification were constructed based on the combinations of test approaches that include NDS and PDS. This model was developed from the relationship between the prevalence and the positive and negative predictive values of the risk factors to estimate the case identification rate and the proportion of mothers indentified for subsequent testing. In the CEA of case detection, the strategy that can detect all patients in a population with low screening costs is the most cost-effective. The costs per case identifications range from £7.64 - £21.12 depending on the screening test used. According to the results, S2 using a PDS approach with selective screening followed by two-step approach using FPG and 75g OGTT yielded the highest rate of case identification and the lowest costs for screening tests for GDM. S2 is a screening test strategy that was published by NICE in 2008 (NICE, 2008b). The NICE guidelines recommend NDS with selective screening that screens all mothers with risk factors at first visit followed by FPG and 75g OGTT.

This model assumed a GDM prevalence of 3.5% in the population and that diagnostic tests have 100% sensitivity and specificity (gold standard). Combining the test results and assuming the diagnostic tests have 100% sensitivity and specificity (gold standard), the probability of TP outcomes in PDS is equal to the prevalence in the population. In PDS, the sensitivity for each strategy was 100% where the probability of case identification equaled the prevalence in the population (TP) and the probability of obtaining FN results equaled to zero. However, in PDS, the probabilities for TP and FN are the same for each strategy, whereas, the probabilities of FP and TN outcomes vary depending on the specificity of the test employed. On the other hand, using this gold standard for tests, the specificity for each strategy in the model in NDS was 100%, and allowed for all patients without disease who have a negative test result to be identified (TN), whereas the probability of obtaining FP results is equal to zero. In NDS, the probabilities for TN and FN are the same for each strategy, whereas, the probabilities of TP and FN outcomes vary depending on the sensitivity of the test employed. There are additional costs associated with the unnecessary testing of FP outcomes in PDS. Unlike PDS there is a high probability of FN outcomes in NDS. Consequently patients are miss-

diagnosed. Using PDS for screening tests, all patients in the population can be detected. The case identification rate is a significant component of the model, as treatment costs and outcomes are predicated in it.

No CEA studies were identified in the review in Chapter 6 that presented outcomes in terms of case detection of GDM. One potentially essential outcome for the detection of GDM is the identification of women who are at high risk of developing type 2 DM or at high risk of GDM in subsequent pregnancies. The model for economic evaluation for GDM should account for the effectiveness of postpartum screening for type 2 DM. However, case identification may not be an appropriate outcome measure to use for CEA, as the objective to introduce a screening programme is to offer treatment and prevent short term and long term complications of GDM. If, however the clinician wishes to consider outcomes in terms of case detection the screening strategy with the highest detection rate is the most cost-effective. However, in PDS, the screening test strategy that has the highest detection rate may produce a higher probability of FP outcomes and so patients without disease receive treatment. Clinicians should be aware of a high number of FP results as well, otherwise patients may receive unnecessary treatment which may be harmful to their health.

10.2.5 Short term complications

This model estimates the cost-effectiveness of screening for and treating GDM, considering short term complications in both mothers (pregnancy period plus 3 months after delivery) and offspring (neonatal period plus 11monthsof infancy) over a period of one year. The combinations of test approaches NDS and the PDS approach were used to construct the decision trees. The cost effectiveness acceptability curve (CEAC) demonstrated that a willingness to pay at a threshold value of £30,000/QALY, the probability that S1 (NDS) is cost-effective was 64%, and when compared with other screening test strategy this is the most effective strategy at this threshold. From this it is concluded that S1 (NDS) is a cost-effective screening test strategy to prevent short term complications in both mothers and offspring, in comparison with the other screening strategies. The combination of a screening test (random plasma glucose) followed by a diagnostic test (75g OGTT) and screening with the highest case identification rate is not necessarily the most cost-effective strategy with respect to the prevention of short term complications. It is potentially worthwhile undertaking further research to explore screening tests for GDM if further research costs less than £50,000. If it costs more than £50,000, further research is not worthwhile.

In short term complications NDS was found to dominate PDS for screening tests. With the same GDM prevalence of 3.5% in the population and assuming diagnostic tests have 100% sensitivity and specificity (gold standard), the probability of TP outcomes in PDS is equal to prevalence in the population (3.5%). In PDS, the probabilities for TP and FN are the same in each strategy, whereas, the probabilities of FP and TN outcomes vary depending on the specificity of the test employed. A consequence of screening in PDS is the unnecessary inconvenience and worry associated with FP, such as side effects of treatment care during the gestational period. The detrimental effect of treating mothers without the disease changes utility outcomes in NDS and PDS to approximately the same extent. Treatment for GDM during gestation has a low impact on utility outcomes; therefore FP cases do not influence PDS results for short term complication. However, as there is a higher probability of FP outcomes in PDS, for diseases whose treatment results in serious side effects, PDS should not be considered as a screening test strategy.

On the other hand in NDS, the probabilities for TN and FN are the same in each strategy, whereas, the probabilities of TP and FN outcomes vary depending on the sensitivity of the test employed. Unlike PDS there is a high probability of FN outcomes in NDS. Consequently patients are miss diagnosed, and so patients with the disease do not receive treatment leading to high costs associated with the treatment of serious complications and low health-related quality of life in short term complications. Therefore, the probability of FN or misdiagnosis in NDS have driven the cost-effective outcomes in short term complications and the purpose of introducing screening tests is to prevent short term complications in both mothers and offspring.

The RPG test has reasonable sensitivity with low levels of prevalence in the population. This means that the additional number of GDM cases that would be missed is small. It is also represents a lower cost strategy than testing all women with 75g OGTT. Assuming 100% specificity and sensitivity, using 75g OGTT alone no longer appears to be a cost-effective strategy. Therefore, confirming an RPG positive with 75g OGTT is still cost-effective as the cost of the test is offset completely not only the savings made from not treating FP cases but also saving costs of treatment in short term complication in both mother and offspring.

10.2.6 Long term complications

Case detection and short term complications are not the endpoint outcomes for GDM screening tests. This economic evaluation of screening tests for GDM estimates cost-effectiveness considering both short term complications (QALY over one year) and long

term complications together (lifetime QALY of type 2 DM). According to this model, at a willingness to pay threshold value of £30,000/QALY, the estimated probability that S4 (PDS) is cost-effective is 42%, the highest compared with other screening strategies at this threshold. S4 (PDS) is a two step approach with selective screening that involves screening for GDM with risk factors at initial visit followed by 75g OGTT. With such a high population EVPI, it can be concluded there is potential in undertaking future research to reduce the uncertainty that surrounds the question of screening tests for GDM using S4 (selective screening).

Considering both short term and long term complications together in order to reduce risk of type 2 DM, PDS was found to dominate NDS for screening tests. Using the gold standard for tests, the sensitivity and specificity for diagnostic tests in the model were 100%, PDS has a higher proportion of patients in the population with positive screening test results (TP). In term of short term complications, patients with disease that have also been diagnosed with the disease receive treatment during gestation, which reduces costs associated with the treatment of complications in both mothers and offspring. There is a higher probability of FP outcomes in PDS, and so a high proportion of patients without disease subsequently undergo unnecessary treatment (FP) for short term complications, with the possibility of an associated high level of unnecessary costs. Treatment given to mothers without disease during gestation for GDM has a low impact on utility outcomes. In long term complications, in PDS, mothers who have a history of GDM in both the TP and FP outcome groups undergo postpartum screening tests for type 2 DM three times over a 10 year period. Patients with type 2 DM that are diagnosed early receive treatment to prevent type 2 DM with complications. The treatment will result in improved quality of life and cost savings, as the patients will not need to be treated for type 2 DM with complications. Therefore, the probability of FP or over diagnosis in PDS has driven the cost-effective outcomes in long term complications.

Test errors result in FN outcomes in NDS, whereas the probability of obtaining FP results is equal to zero in NDS. Mothers in the FN group have a high risk of developing type 2 DM, however, as mothers with symptoms of type 2 DM in the FN group will not be tested for type 2 DM, they will not be detected or treated. As disease in these patients is not detected or treated they have an increased risk of developing complications of type 2 DM. FN or miss diagnosis outcomes in NDS did not influence the results of long term complications. Screening with the highest case identification rate is the most cost-effective strategy for long term adverse complications. The detection rate is an important component of screening tests when considering long term complications, as treatment costs and the effect of treatment are dependent on it.

This model assumes that long term complications and short term complications are conditionally independent of screening tests for GDM. Adding costs and outcome results for both short term complications and long term complications together alters the ranking of screening test strategies according to their cost effectiveness, as shown in table 9.4 For example, for short term complications, S1 (NDS) is ranked second according to costs, whereas in long term complications S1 (NDS) is ranked fourth, however when both are added together S1 (NDS) is ranked sixth compared to other strategies. Therefore, the construction of a model for short and long term complications that are conditionally independent on screening tests, using PDS and NDS impacts the cost-effectiveness of the strategies.

10.2.7 Impact of uncertainty

The aim of probabilistic modeling is to reflect the uncertainty in the input parameters of the decision model and describe what this means for uncertainty over the output of interest: the measure of cost, effect and cost-effectiveness. Considering the deterministic model, costs and effectiveness in both short and long term complications are seen to be slightly different for each strategy. However, the mean costs and effectiveness of the probabilistic model show similarity to the deterministic model. The deterministic results showed that screening for GDM was cost-effective. However, this result was sensitive to the addition of uncertainty. At willingness to pay threshold of £30,000/QALY, the probabilistic model showed that the screening program was cost-effective only 64% of the time in short term compilations and 42% of the time in long term complications. Thus it can be inferred that the opportunity loss of choosing the screening strategy is high at a willingness threshold of £30,000/QALY. In the populations EVPI at the same willingness to pay, results show that it is worthwhile collecting more information to explore screening tests for GDM in both short term complications (£494,990) and long term complications (£107 million). EVPPI in long term complications approximately £784,042, it can be concluded there is potential in undertaking future research to reduce the uncertainty that surrounds the question of screening tests for GDM. Clearly it is the parameters for long term complications that are driving uncertainty in the model, and leading to uncertainty in the cost-effectiveness decision. In addition, one-way sensitivity analysis was performed to further examine the outcomes that had an impact on the cost-effectiveness of screening for GDM. Varying the values of the prevalence parameters slightly changed the results of the incremental cost effectiveness ratio.

10.3 Policy implications

Healthcare resources are limited and as a result their use must be as effective as possible, an issue that those working in the health care system have always been acutely aware of. Consequently decisions made by public health officials and clinicians, that relate to the use of resources, must not only take into consideration the effectiveness of procedures but also the costs that they incur. In this respect, cost-effectiveness analysis provides a useful method to make such decisions more scientific in nature.

Economic assessments of diagnostic tests are inherently difficult as assessments of therapeutic intervention, mainly because of uncertainty about the relation between diagnosis and outcomes. The outcomes of screening tests are more usually measured by additional case detection which depends on the sensitivity and specificity of the tests. The trade-off between sensitivity and specificity or detection and unnecessary testing is a central issue for screening and diagnostic tests and the accuracy of the tests, in terms of sensitivity and specificity, is a consideration of great importance for clinicians. Even if clinicians are able to keep both specificity and sensitivity high in tests, FP and FN outcomes will still persist. However, on combining the tests, when the second test is performed to confirm the first test, there is a trade-off between the sensitivity and specificity of the tests. The two-step approach is considered to be better than using individual tests. If the first test has high specificity and low sensitivity, then by using PDS the second test can increase sensitivity. In contrast, if the first test has high sensitivity and low specificity, by using NDS the second test can increase specificity.

In PDS, it is patients that are negative in the first test (screening test) that undergo an additional second test (diagnostic test).There are a higher proportion of patients in the population with positive screening test results in this strategy (TP), whereas there is only one possible outcome for FN results in PDS. On the other hand, patients without disease that test positive in either the first or second test identify as FP. TP and FP results are the main outcomes of PDS where screening and diagnostic tests involve positive results. In PDS, clinicians should be concerned about tests with high specificity, as tests that can identify patients without the disease who have negative results reduce the numbers of FP patients.

In NDS, only patients that test positive in the first test (screening test) undergo a second test (diagnostic test), however only a few people test positive in the second test and so only a small proportion of the population with the disease are detected (TP). The

only possible TP outcome in NDS is for diseased patients that test positive in both screening and diagnostic tests. Additionally, in NDS diseased patients that have negative test results in either the screening or diagnostic tests identify as false negative (FN), whereas patients without disease that have positive results in both the screening and diagnostic tests identify as false positive (FP), the only possible FP outcome in NDS. TN and FN results are therefore the main outcomes of NDS where either screening or diagnostic tests involve negative results. Therefore there is a high probability that patients without the disease have a negative result in NDS. In NDS, clinicians should be concerned about tests with high sensitivity, as tests that can identify patients with the disease who have positive test results reduce the numbers of FN patients.

An earlier diagnosis can be the result of screening; however it has been shown that screening tests do not always result in benefits to the patients that are screened. Some possible adverse effects that can result from screening and diagnosis include misdiagnosis, over-diagnosis, and creating a false sense of security (Laking et al., 2006). This economic evaluation of screening for GDM was applied to both NDS and PDS approaches in order to help inform the clinician and policy maker of possible alternative outcomes. This allows the clinician to consider the accuracy of the test results in terms of TP, FN, FP and TN. The possible adverse effects that can result from misdiagnosis (FN) are worse than the adverse effects that can result from over-diagnosis (FP) cases. Due to errors in the results (FN, FP), it can be understood why PDS replaces NDS as the most cost effective strategy when considering long term complications as opposed to short term complications. The unnecessary costs that result from FN screening tests and of treatment due to FP results are highlighted as an issue of error in the accuracy of the tests. However, all guidelines for GDM screening tests are performed as NDS with regard to both screening tests and postpartum screening. This study found that long term complications are the most important factor affecting the results of screening test strategies. Therefore, policy makers or clinicians should consider screening tests for GDM in order to prevent long term complications and perform screening tests as PDS.

Accordingly, the results of this study provide useful information for policymakers deciding whether to adopt screening for GDM with treatment followed by screening for type 2 DM during postpartum to delay or prevent onset of type 2 DM. This economic evaluation of screening for GDM was applied to both of these strategic approaches in order to help inform the clinician and policy maker of possible alternative outcomes. In terms of screening tests, numerous national guidelines recommend screening tests for GDM, however only a few guidelines recommend postpartum screening tests for type 2

DM among women with histories of GDM. If women have received previous diagnoses of GDM, this should trigger regular screening for type 2 DM so that it is discovered early on, before the onset of symptoms or the development of complications associated with type 2 DM. The early detection and prevention of type 2 DM should potentially result in cost savings and clinical benefits. For mothers with GDM that have been correctly diagnosed (TP) and mothers diagnosed with GDM that do not have disease (FP), postpartum screening tests for type 2 DM should be employed to allow early detected and delayed progression of type 2 DM with complications. Given the link between GDM and the highly elevated risk of future type 2 DM, and a rapidly increasing number of type 2 DM, policy makers should consider GDM in relation to type 2 DM in both mothers and offspring. As a consequence of this study, policy makers should consider the implementation of screening tests for type 2 DM in the general population too, in order to prevent the development of type 2 DM with complications.

Although there is no universal standard of screening and care for GDM, most high-income countries have developed policies or guidelines for GDM in terms of screening tests and the treatment of GDM (Jiwani et al., 2012). National guidelines for screening tests and the treatment of GDM can inform the policy maker but need to be tailored to local prevalence as well as the structure and capabilities of the local health care system in order to deliver results in a way that is meaningful to local policy makers.

Moreover, VOI is a useful tool for decision makers to evaluate screening tests for GDM. Using this tool decision makers are able to make informed decisions with regard to the adoption or rejection of new interventions, based on current evidence, or indeed whether additional information should be obtained to help decision making.

10.4 Limitations

In any analytical model there are limitations that affect the interpretation of its findings, as is the case in this study.

The most cost-effective screening test strategy might vary according to local prevalence. This model acknowledges that context specific prevalence data is important and this thesis conducted a systematic review to identify prevalence estimates. No estimate of the Scottish population was identified. In the absence of this specific parameter, this thesis used a broadly applicable estimate from NICE of 3.5%, based on an expert suggestion. The NICE estimate represents the best available evidence. This thesis acknowledges that this is a limitation of the model and decision makers should look to conduct a study to estimate GDM prevalence to reduce this uncertainty.

However, this thesis felt the uncertainty in this estimate was well represented as this model explored a range of values in the probabilistic model as well as different one-way scenario analyses. In terms of representing the Scottish context, this thesis, therefore takes into account some important parameters from the Scottish context to conducted an economic evaluation for screening tests for GDM, such as the EQ-5D mean weighted health state index by age and standard region (Scotland), the incident rate of type 2 DM reported by NHS Scotland, and the annual pregnancy population to determine the population value for EVPI.

Moreover, many screening test strategy models included GDM risk factors, either alone or in combination. The model does not capture the impact and interdependence of multiple risk factors. As the proportion of the population with risk factors various for mothers with GDM, there would be concomitant increase in the proportion with multiple risk factors, which would change the positive and negative value of test results. However, there is no reason why the proportion of the population with multiple risk factors would constant with respect to prevalence.

Decision analysis has its own specific limitations in representing outcomes in the real-world. During pregnancy, GDM women receive interventions including dietary control, home glucose monitoring, as well as insulin therapy when needed. The treatment care for GDM in this study included dietary control, home glucose monitoring and insulin therapy. This study assumes that all women identified as having GDM received different treatment, that they received this treatment without fail and that all treatment was completed. In a real word situation, if initial treatment of dietary control during gestation fails, patients would be offered alternative forms of treatment, such as insulin therapy. However, this model did not account for the effect of treatment during gestation and treatment to prevent type 2 DM.

Short term outcomes in both mothers and offspring are modeled as separate outcomes that cannot coexist; for example an offspring in this model cannot have both macrosomia and shoulder dystocia, although this would be possible in the real world. The coexistence of short term outcomes may result in higher costs and lower QALY than for individual adverse outcomes for both mothers and offspring.

Patients with GDM may suffer from the side effects of treatment or they may for example spend a significant amount of time searching for information about the condition of their health, despite this disutility in mothers associated with stress of a GDM diagnosis or treatment during the gestation period was not taken into account in the model. However, one study on health-related quality of life related to the effect of

treatment of GDM on pregnancy outcomes found that QALY was not negatively impacted by a GDM diagnosis (Crowther et al., 2005).

In this model, screening test accuracy information uses the same test threshold to determine a positive result for fasting plasma glucose and random plasma glucose as other studies in the literature. This model assumed the second test to have 100% sensitivity and 100% specificity. The model treated sensitivity and specificity independently and there was no trade-off between sensitivity and specificity with respect to the accuracy of the tests. This is a limitation of the model. However, independent modeling of sensitivity and specificity has been shown to underestimate true test performance when studies use the same and different thresholds. In addition, if a probabilistic modeling approach is used, the approach will estimate the uncertainty incorrectly(Novielli et al., 2010). The probabilistic approach randomly sampled sensitivity and specificity independently; thus it was possible that a very high sensitivity and low specificity were simultaneously sampled. Consequently, the true cost-effectiveness of the test has not been ascertained however, as the optimal operating point could not be identified, as only one test threshold has been used (Sanghera et al., 2013).

Sometimes a test result can be positive in patients who do not actually have the disease, which is called serendipity. Serendipity refers to the coincidental detection of disease following the review of a false-positive screening test result. As the test result is abnormal due to factors other than the presence of the target disease, disease detection is not attributed to the screening test (Vis et al., 2002). Serendipity in this model refers to identifying cases (TP) in the absence of a screening test. Even without a screening test for GDM, mothers might report to a clinic with a symptom of GDM. Although S5 includes no screening tests and therefore no case detection, it might therefore underestimate both the costs and outcomes of the tests.

10.5 Generalisability

To assess the relative generalisability of the decision model some limitations need to considered. Firstly, the data used in this analysis were derived from the best available data and the assumptions used in the model were designed before the analysis. The construction of the decision tree for the economic evaluation model for GDM screening tests begins with true disease status. The systematic review in chapter 5 stated that the variation in the prevalence of GDM can be explained by screening tests and population characteristics, therefore, in this model, the prevalence for GDM of 3.5% reported by NICE is used to represent the baseline parameter for true disease status. The NICE report of prevalence was chosen because its data is based on that from England &

Wales, which has a similar population demographic to Scotland. The prevalence of GDM at 3.5% represents the prevalence of GDM in countries where the majority of the population is white. This represents a potential limitation in the generalisability of the findings because the prevalence of GDM varies dependent on ethnicity, which is an important risk factor of GDM. Geographical characteristics can affect the incidence and prevalence of the disease and lead to generalization of cost-effectiveness results across geographic areas. Moreover, it is important to consider heterogeneity of the disease frequency in the population, as this may introduce bias. In addition, this model used multi-parameter evidence synthesis derived from systematic review and Meta-analysis. For example, clinical effectiveness data for short term complications in mothers and offspring were derived from meta-analysis, whereas clinical effectiveness data for long term complications were derived from systematic review.

Three independent decision trees were constructed, including screening tests for GDM, short term complications and long term complications, covering all outcomes of screening tests for GDM. Moreover, this model assumes that short term complications and long term complications are conditionally independent of screening tests for GDM, which allows policy makers to focus on each part of the model separately. Conditionally independent decision trees for screening tests for GDM form a more robust decision tree in terms of diagnostic test evaluation. The four screening tests in this model (other than 'no screening') include either a one step approach (the diagnostic test is performed without prior screening) or a two step approach (diagnostic test is performed to confirm a positive screening test) and a universal or selective approach, which together cover all the fundamental approaches of performing screening tests, which allows for generalisability.

Lastly, this study performed a decision analysis from a health care provider's perspective, not a societal perspective, in order to inform policy makers in Scotland. This model excluded productivity costs or out-of-pocket direct costs incurred by the patient. Therefore, the perspective of the study (payer or societal) can limit generalisability.

10.6 Future research

Evidence in this thesis and from previous research has shown that screening tests for GDM are cost-effective. This is the first cost-effectiveness study that used Scottish screening test guidelines based on the UK population prevalence. However, further important research questions remain, as outlined below.

10.6.1 Expected value of sample information (EVSI)

The population EVPI at a threshold value of £30,000/QALY was £124 million, indicating that commissioning further research is of extremely high value for long term complications. Having determined this and having used EVPPI to specify the most likely type of information and type of study that is potentially required. Long term adverse complication parameters are clearly what are driving uncertainty in the model and are leading to uncertainty in the cost-effectiveness decision. According to the EVPPI results, for particular research designs, an expected value of sample information (EVSI) analysis can be performed to assess the additional benefit that can be expected from increased sample size before a decision is made. The additional information that can be acquired from the sample can allow for more informed and in turn better decisions. EVSI can estimate what this expected improvement may be before sample data is examined, and therefore allows an evaluation of the costs involved of the additional research.

10.6.2 Developing an economic evaluation of screening tests for type 2 DM after diagnosis of GDM

Further research on the cost-effectiveness of type 2 DM after postpartum in women with GDM should be considered. Also, this future research could examine the optimum number of times that screening should be performed and the corresponding time frame for type 2 DM. One potentially important outcome for the detection of GDM is identification of mothers who are at high risk of subsequent type 2 DM or at high risk of GDM in subsequent pregnancies. In mothers with GDM, shortly after delivery, glucose homoeostasis is restored to non-pregnancy levels, but affected women remain at high risk of developing type 2 DM in the future. Screening tests and treatment after postpartum intervention are highly effective in reducing the incidence of type 2 DM. It is recommended that women who have had GDM have their glucose tolerance assessed 6 weeks after delivery.

There are many guidelines that suggest postpartum follow up glucose tolerance tests for mothers with GDM (WHO, 1999a) (ACOG, 2000) (American Diabetes, 2004) (NICE, 2008d). Those guidelines suggest screening all postpartum mothers with GDM over different periods of time and with different screening test techniques. Postpartum screening for type 2 DM is a controversial issue in terms of the appropriate time and methods for screening (Bentley-Lewis et al., 2008)

10.6.3 Long term study to investigate the potential link between GDM and offspring outcome

The population EVPI at a threshold value of £30,000/QALY was £142 million, and therefore research that costs less than this value is potentially worthwhile. However, the EVPI for the individual parameters and sub-group parameters for long-term adverse complications, the EVPPI analyses show that there is zero value in undertaking further research. However, the EVPPI result from the group of long-term adverse complications parameters was £832,423 at the population level and suggests that further research would be of worthwhile potential. It was difficult to identify which individual parameters were driving uncertainties in the cost-effectiveness decision; therefore, the type of further research that would be required is likely to be a randomised trial. Such a trail should be long-term trial in nature in order to study type 2 DM among women with histories of GDM and offspring born to women with GDM. According to the VOI analysis, the EVPI and EVPPI provided information about whether it was worthwhile to conduct further research. It is clear that the parameter for long term complications is driving uncertainty in the model. Given the result of the EVPPI, the type of further research that would be required is likely to be a trial to evaluate the development of type 2 DM in mothers with GDM and in offspring born to mothers with GDM. However, there were no studies in the review that looked into the effect of treatment in women with GDM and offspring born to mothers with GDM. Most studies looked at the short term effects of treating GDM compared to no treatment, such as the body mass index or obesity in adults. The emerging field of developmental origins of human disease suggests that glucose intolerance during pregnancy is associated with long-term outcomes such as type 2 DM, hypertension and obesity.

The application of VOI analysis is useful as it helps decision makers to make informed decision as to whether a new intervention should be (used) or reject based on current evidence. . If funding were made available for further research in this area, a randomised controlled trial (RCT) would provide strong evidence as to whether long term complications in mothers with and without GDM develop type 2 DM. Therefore, long term study to investigate the potential link between GDM adult outcomes for the offspring of mothers with GDM is needed. However, long-term trials cost more in terms of money and time. An alternative study design such as retrospective data base analysis would be better to estimate long term complications.

10.7 Conclusion

This thesis discussed and demonstrated the application of decision analytic modelling in healthcare. Both NDS and PDS have been included in this economic evaluation of screening tests for GDM in order to inform the clinician and policy maker. The combinations of the tests in NDS and PDS are alternative approaches in evaluating the economic assessments of diagnostic tests. Screening test strategies for GDM and screening tests for other diseases usually recommend NDS to screen and detect disease in individuals in the general population. The NDS and PDS approaches allow the clinician to consider test results in terms of the differences in FN and FP test results. Due to these errors in the tests there is a higher probability of FN outcomes in NDS and so patients with a true disease state do not receive treatment. In contrast, there is a higher probability of FP outcomes in PDS and so patients without disease receive treatment.

Three independent decision trees were constructed, including screening tests for GDM, short term complications and long term complications, which allows policy makers to focus on each part of the model separately. Economic modeling in this study shows that selective screening tests with 75g OGTT are cost-effective when compared to other strategies. Selective screening generated more QALYs at a lower cost than other screening test strategies. This study also found that the cost effectiveness of screening tests for GDM to prevent short term complications is dependent on the probability of GDM being left undiagnosed based on strategy employed. Additionally, treatments during gestation are important as they reduce additional cost that may be required to treat serious adverse complications.

In long term complications, screening tests for GDM that can detect all patients in the population or result in over-diagnosis are cost-effective. If mothers have received previous diagnoses of GDM, regular screening should be performed for type 2 DM so that it is discovered early on, before the onset of symptoms or the development of complications associated with type 2 DM. Screening for GDM and its subsequent treatment presents an important opportunity to reduce type 2 DM. The outcome of screening tests is dependent on the test's sensitivity and specificity. NDS and PDS allow the clinician to understand error in screening test results (FN, FP).

Modelling procedures, in which intermediate measures that have a connection to the assessments of diagnostic tests are converted into costs and long-term health benefits, are usually needed in order to calculate the cost-effectiveness of diagnostic testing. The VOI method makes meaningful recommendations to decision makers, who

can then make informed decisions as to whether a new intervention should be adopted or rejected based on current evidence, or whether further information is required to help make the decision, as opposed to making decisions on subjective reasoning.

Appendix I: Literature search strategy

All concept of gestational diabetes mellitus (Chapter 4)

1. exp Pregnancy/
2. exp Diabetes Mellitus, Type 1/ or exp Diabetes Mellitus, Lipoatrophic/ or exp Diabetes Mellitus, Type 2/ or exp Diabetes Mellitus/
3. 1 and 2
4. exp Diabetes, Gestational/
5. mass screening/
6. (screen* or test* or detect* or (case adj2 find*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
7. 5 or 6
8. (pregnan* adj12 diabetes*).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
9. 3 or 4 or 8
10. 7 and 9
11. (guideline$ or guidance).tw.
12. (management and treatment).tw.
13. 11 or 12
14. 10 and 13
15. Pregnancy Complications/
16. ((adverse or negative) adj3 outcome$).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
17. 15 or 16
18. 10 and 17
19. ("true positive" or "false positive" or "false negative" or "true negative" or "predictive value$" or ROC or likelihood or evaluat$ or accura$ or "receiver operating characteristic").mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
20. ((sensitivity and specificity) adj3 ("glucose challenge test" or "fasting plasma glucose" or "random plasma glucose" or "glucose tolerance test$" or GCT or FPG or RPG or OGTT or "glucose screening test$")).tw.
21. ("glucose challenge test" or "fasting plasma glucose" or "random plasma glucose" or "glucose tolerance test$" or GCT or FPG or RPG or OGTT or "glucose screening test$").tw.
22. 19 and 20 and 21
23. 10 and 22
24. Quality of life/
25. ("qualtiy adjusted life years" or "disabiltiy adjusted life years" or QALY or QALYs or DALY or DALYs).mp.
26. 24 or 25
27. 10 and 26
28. 14 or 18 or 23 or 27
29. remove duplicates from 28

Prevalence (Chapter 5)

1. exp Pregnancy/
2. exp Diabetes Mellitus, Type 1/ or exp Diabetes Mellitus, Lipoatrophic/ or exp Diabetes Mellitus, Type 2/ or exp Diabetes Mellitus/
3. 1 and 2
4. exp Diabetes, Gestational/
5. mass screening/
6. (screen* or test* or detect* or (case adj2 find*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
7. 5 or 6

8. (pregnan* adj12 diabetes*).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
9. 3 or 4 or 8
10. 7 and 9
11. exp Epidemiology/
12. epidemiolog*.mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
13. exp Prevalence/
14. prevalen*.mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
15. 11 or 12 or 13 or 14
16. 9 and 15

Economic evaluations of screening test for GDM (Chapter 6)
MEDLINE (Ovid) (monthly search)

1. economics/
2. exp "costs and cost analysis"/
3. Economics, Dental/
4. exp "economics, hospital"/
5. Economics, Medical/
6. Economics/ or Economics, Nursing/
7. Economics, Pharmaceutical/ or Economics/
8. (economic$ or cost or costs or costly or costing or price or prices or pricing orpharmacoeconomic$).ti,ab.
9. (expenditure$ not energy).ti,ab.
10. value for money.ti,ab.
11. budget$.ti,ab.
12. or/1-11
13. ((energy or oxygen) adj cost).ti,ab.
14. (metabolic adj cost).ti,ab.
15. ((energy or oxygen) adj expenditure).ti,ab.
16. or/13-15
17. 12 not 16
18. letter.pt.
19. editorial.pt.
20. historical article.pt.
21. or/18-20
22. 17 not 21
23. Animals/
24. Humans/
25. 23 not (23 and 24)
26. 22 not 25
27. exp Pregnancy/
28. exp Diabetes Mellitus, Type 1/ or exp Diabetes Mellitus, Lipoatrophic/ or exp Diabetes Mellitus, Type 2/ or exp Diabetes Mellitus/
29. 27 and 28
30. exp Diabetes, Gestational/
31. mass screening/
32. (screen* or test* or detect* or (case adj2 find*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
33. 31 or 32
34. (pregnan* adj12 diabetes*).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
35. 29 or 30 or 34
36. 33 and 35
37. 26 and 36
38. remove duplicates from 37

EMBASE (Ovid) (weekly search)

1. health-economics/
2. exp economic-evaluation/
3. exp health-care-cost/
4. exp phamacoeconomics
5. 1 or 2 or 3 or 4
6. (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab
7. (expenditure$ not energy).ti,ab
8. (value adj2 money).ti,ab)
9. budget$.ti,ab
10. 6 or 7 or 8 or 9
11. 5 or 10
12. letter.pt.
13. editorial.pt.
14. note.pt
15. 12 or 13 or 14
16. 11 not 15
17. (metabolic adj cost).ti,ab.
18. ((energy or oxygen) adj cost).ti,ab.
19. ((energy or oxygen) near expenditure).ti,ab.
20. 17 or 18 or 19
21. 16 not 20
22. exp animals/
23. exp animal-experiment/
24. nonhuman/
25. (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh
26. 22 or 23 or 24 or 25
27. exp human/
28. exp animal-experiment/
29. 27 or 28
30. 26 not (26 and 29)
31. 21 not 30
32. exp Pregnancy/
33. exp Diabetes Mellitus, Type 1/ or exp Diabetes Mellitus, Lipoatrophic/ or exp Diabetes Mellitus, Type 2/ or exp Diabetes Mellitus/
34. 27 and 28
35. exp Diabetes, Gestational/
36. mass screening/
37. (screen* or test* or detect* or (case adj2 find*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
38. 31 or 32
39. (pregnan* adj12 diabetes*).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sh, tn, dm, mf, dv, kw]
40. 29 or 30 or 34
41. 33 and 35
42. 26 and 36
43. remove duplicates from 42

Appendix II: Guideline of screening test for GDM and reference

Organisation	Year	Diagnostic method	Threshold values (mmol/l)				Type of screening	Remark
			0h	1h	2h	3h		
O'Sullivan-Mahan(Carpenter and Coustan, 1982) (O'Sullivan and Mahan, 1964)	1964	100g OGTT	≥5.2	≥9.9	≥8.5	≥7.7	Universal	
NDDG (National Diabetes Data Group, 1979) (Turok et al., 2003)	1979	50g GCT		≥7.8			N/A	Two or more value must be met
		100g OGTT	≥5.8	≥10.5	≥9.2	≥8.1		or exceeded
C&C(Coustan and Carpenter, 1998)	1982	50g GCT		≥7.8			Selective	Age > 24 years
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		
EADS(Lind and Phillips, 1991)	1991	FPG	≥4.8				Universal	
		50g GCT		≥8.2				
		75g OGTT	≥6.0		≥9.0			
Thirst international Workshop (Metzger and Coustan, 1998)	1991	50g GCT		≥7.8			Selective	
		75g OGTT	≥5.3	≥10.0	≥8.6			
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		
Fourth international Workshop (Metzger et al., 2007)	1998	50g GCT		≥7.8			Selective	Screening test : optional,
		75g OGTT	≥5.3	≥10.0	≥8.6			2 step 50 /100 g or 1
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		step 75g GCT Diagnostic test : 75 g or 100g
								Two or more abnormal values
ADA(Jovanovic, 1998) (American Diabetes Association, 1998)	1998	50g GCT		≥7.2			Selective	Obesity, > 25 years of age,
		75g OGTT	≥5.3	≥10.0	≥8.6			Family history of 1st relative DM, Race(Hispanic ,Asian,
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		African, Pacific Islander)
								Two or more abnormal values and One or two step approach
WHO(WHO, 1999a)	1998	50g GCT		≥7.8			Universal	One-step approach to screening
		75g OGTT	≥7.0		≥7.8			and diagnosis, one or more criteria must be met or exceeded and one step approach
CDA(Metzger and Coustan, 1998)	1998	50g GCT		≥7.8			Selective	Or C&C conversion and
		75g OGTT	≥5.3	≥10.6	≥8.9			two step approach
ADIPS(Hoffman et al., 1998)	1998	50g GCT		≥7.8			Universal	Universal unless low GDM
		75g OGTT	≥5.5		≥8.0 (Aus) ≥9.0 (NZ)			Incidence or resources limited
ACOG(Serlin and Lash, 2009)	2001	50g GCT		≥7.8			Universal /	Test every one or in
		75g OGTT	≥7.2		≥7.8		selective	case of risk factors, history of
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		previous GDM
SIGN(SIGN, 2001)	2001	RPG	≥5.5				Universal	Urine for screening for every
		FPG	≥5.5					visit
		75g OGTT	≥7.0		≥9.0			
CREST(Clinical Resource Efficiency Support Team, 2001)	2001	RPG	≥5.5				Universal	Urine for screening for every
		75g OGTT	≥6.0		≥9.0			visit and 28 weeks gestation
SOGC(Berger et al., 2002)	2002	50 g GCT		≥7.8			Universal	Routine screening with 50 g
		75g OGTT	≥5.3	≥10.0	≥8.6			GCT except in mother who fulfill
		100g OGTT	≥5.3	≥10.0	≥8.6	≥7.8		The criteria for low risk

HTA UK(Scott et al., 2002)	2002	50g GCT		≥7.8		Selective	Very selective screening based on age, obesity and ethnic origin
NICE(NICE, 2008d)	2008	FPG	≥7.0			Selective	Women with on of risk factor
		75g OGTT	≥7.0		≥7.8		should be offer screening test
IADPSG(Metzger et al., 2010a)	2010	FPG	≥5.1	≥10.0	≥8.5	Universal	
		75g OGTT	≥7.0		≥7.8		
Consensus(Moses, 2010)	2010	75g OGTT	≥7.0		≥7.8	Universal	
SIGN(SIGN, 2010)	2010	FPG	≥5.1			Selective	one or more value must
		75g OGTT	≥10.0		≥8.5		be met or exceeded

Organisation (C&C) Indicated Carpenter and Constant, (NDDG) National Diabetes Data Group, (WHO) World Health Organisation, (ADIPS) The Australasian Diabetes in Pregnancy Society,(ADA) American Diabetes Association, (CDA)Canadian Diabetes Association, (ACOG) American College of Obstetricians and Gynecologists ,(EASD) European Association for the Study of Diabetes. (CREST) linical Resource Efficiency Support Team, (SIGN) The Scottish Intercollegiate Guidelines Network, (SOGC) The Society of Obstetricians and Gynaecologists of Canada (NICE) The National Institute for Health and Clinical Excellence ,(IADPSG) International Association of Diabetes and Pregnancy Study Groups Consensus Panel,(Aus) Australia, (NZ) New Zealand : **Tests** (OGTT) Oral Glucose tolerance Test, (GCT) Glucose challenge test, (FPG) Fasting Plasma Glucose, (RPG) Random Plasma Glucose : **Threshold** Now, if you want to convert mg/dl of glucose to mmol/l, you can divide the result in mg/dl by 18 or multiply by 0.055 but, if you want to convert mmol/l of glucose to mg/dl, just multiply by 18.

1. Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. Am J Obstet Gynecol. 1982 Dec 1;144(7):768-73. PubMed PMID: 7148898. Epub 1982/12/01. eng.
2. O'Sullivan JB, Mahan CM. Criteria for the Oral Glucose Tolerance Test in Pregnancy. Diabetes. 1964 May-Jun;13:278-85. PubMed PMID: 14166677. Epub 1964/05/01. eng.
3. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National Diabetes Data Group. Diabetes. 1979 Dec;28(12):1039-57. PubMed PMID: 510803. Epub 1979/12/01. eng.
4. Turok DK, Ratcliffe SD, Baxley EG. Management of gestational diabetes mellitus. Am Fam Physician. 2003 Nov 1;68(9):1767-72. PubMed PMID: 14620596. Epub 2003/11/19. eng.
5. Coustan DR, Carpenter MW. The diagnosis of gestational diabetes. Diabetes Care. 1998 Aug;21 Suppl 2:B5-8. PubMed PMID: 9704220. Epub 1998/08/15. eng.
6. Lind T, Phillips PR. Influence of pregnancy on the 75-g OGTT. A prospective multicenter study. The Diabetic Pregnancy Study Group of the European Association for the Study of Diabetes. Diabetes. 1991 Dec;40 Suppl 2:8-13. PubMed PMID: 1748272. Epub 1991/12/01. eng.
7. Metzger BE, Coustan DR. Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee. Diabetes Care. 1998 Aug;21 Suppl 2:B161-7. PubMed PMID: 9704245. Epub 1998/08/15. eng.
8. Metzger BE, Buchanan TA, Coustan DR, de Leiva A, Dunger DB, Hadden DR, et al. Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus. Diabetes Care. 2007 Jul;30 Suppl 2:S251-60. PubMed PMID: 17596481. Epub 2008/02/27. eng.
9. Jovanovic L. American Diabetes Association's Fourth International Workshop-Conference on Gestational Diabetes Mellitus: summary and discussion. Therapeutic interventions. Diabetes Care. 1998 Aug;21 Suppl 2:B131-7. PubMed PMID: 9704240. Epub 1998/08/15. eng.
10. American Diabetes Association. Gestational diabetes mellitus. Diabetes care. 1998;22 Suppl 21:60-2.
11. WHO. Difinition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Geneva: WHO; 1999 [cited 2011 23 July]. Available from: http://whqlibdoc.who.int/hq/1999/WHO_NCD_NCS_99.2.pdf.
12. Hoffman L, Nolan C, Wilson JD, Oats JJ, Simmons D. Gestational diabetes mellitus--management guidelines. The Australasian Diabetes in Pregnancy Society. Med J Aust. 1998 Jul 20;169(2):93-7. PubMed PMID: 9700346. Epub 1998/08/13. eng.
13. Serlin DC, Lash RW. Diagnosis and management of gestational diabetes mellitus. Am Fam Physician. 2009 Jul 1;80(1):57-62. PubMed PMID: 19621846. Epub 2009/07/23. eng.
14. SIGN. Management of diabetes A natioanl clinical guideline. Edingurgh2001.
15. Clinical Resource Efficiency Support Team. Management of diabetes in pregnancy. Belfast2001 [cited 2011 07 04]. Available from: http://www.gain-ni.org/Library/Guidelines/diabetes_main_doc.pdf.

16. Berger H, Crane J, Farine D, Armson A, De La Ronde S, Keenan-Lindsay L, et al. Screening for gestational diabetes mellitus. J Obstet Gynaecol Can. 2002 Nov;24(11):894-912. PubMed PMID: 12417905. Epub 2002/11/06. eng
fre.
17. Scott DA, Loveman E, McIntyre L, Waugh N. Screening for gestational diabetes: a systematic review and economic evaluation. Health Technol Assess. 2002;6(11):1-161. PubMed PMID: 12433317. Epub 2002/11/16. eng.
18. NICE. NICE Clinical guideline 63: Diabetes in pregnancy Management of diabetes and its complications from pre-conception to the postnatal period. London2008.
19. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. Internation Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy. Diabet Care. 2010;33(3):676-82.
20. Moses RG. New consensus criteria for GDM: problem solved or a pandora's box? Diabetes Care. 2010 Mar;33(3):690-1. PubMed PMID: 20190298. Epub 2010/03/02. eng.
21. SIGN. Management of diabetes A national clinical guideline. Edinburgh2010.

Appendix III: the methodological quality assessment

Criteria for prevalence studies

External validity
Source population
(a) Does the method to select and invite participants result in a study population that covers the complete population or random sample?
Description of eligibility criteria
(b) Is the age range specified?
(c) Are inclusion and exclusion criteria specified?
Participants and nonresponders
(d) Is the response rate > 70%, or is the information on nonresponders sufficient to make inference on the representativeness of the study population?
Description of study period
(e) Is the study period specified?
Description of study population
(f) Are important population characteristics specified?
Internal validity
Data collection
(g) Are the data prospectively collected?
Measurement instrument (questionnaire, interview, additional)
(h) Is the strategy of screening test validated?
(i) Is the period covered by the measurement instrument specified?
Definition of disease
(j) Is a definition of the disease stated?
Reported prevalence
(k) Are age-specific and high risks specific prevalence reported?
(l) Are possible correlates of disease reported?
Informativity
(m) Is the method of data collection properly described?
(n) Are the questions and answer possibilities stated?
(o) Are the reported prevalence rates reproducible?

Based on quality assessment list of systematic review of prevalence (Bishop et al., 2010) (Prins et al., 2002).

Criteria for decision analytic models

Dimension of quality		Questions for criteria appraisal
Structure		
S1	Statement of decision problem	Is there a clear statement of the decision problem?
		Is the objective of the evaluation and model specified and consistent with the stated decision problem?
		Is the primary decision maker specified?
S2	Statement of scope/perspective	Is the perspective of the model stated clearly?
		Are the model inputs consistent with the stated perspective?
		Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?
S3	Rationale for structure	Is the structural of the model consistent with a coherent theory of the health condition under evaluation?
		Are the sources of data used to develop the structure of the model specified?
		Are the causal relationships described by the model structure justified appropriately?
S4	Structural assumptions	Are the structural assumptions transparent and justified?
		Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?
S5	Strategies/comparators	Is there a clear definition of the option under evaluation?
		Have all feasible and practical options been evaluated?
		Is there justification for the exclusion of feasible options?
S6	Model type	Is the chosen model type appropriate given the decision problem and specified causal relationship within the model?
S7	Time Horizon	Is the time horizon of the model sufficient to reflect all important difference between options?
		Is the time horizon of the model, the duration of treatment and the duration of treatment effect describe and justified?
S8	Disease states/ pathways	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of the intervention?
S9	Cycle length	Is the cycle length defined and justified in terms of the natural history of disease?
Data		
D1	Data identification	Are the data identification methods transparent and appropriated given the objective of the model?
		Where choices have been made between data sources, are these justified appropriately?
		Has the quality of the data been assessed appropriately?
		Where expert opinion has been used, are the methods described and justified?
D2	Data modelling	Is the data modelling methodology based on justifiable statistical and epidemiological techniques?

Dimension of quality		Questions for criteria appraisal
D2a	Baseline data	Is the choice of baseline data described and justified?
		Are transition probabilities calculated appropriately?
		Has half-cycle correction been applied to both cost and outcome?
		If not, has this omission been justified?
D2b	Treatment effects	If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?
		Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?
		Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?
		Have alternative assumptions regarding the continuing effect of treatment been explored through sensitivity analysis?
D2c	Costs	Are the costs incorporated into the model justified?
		Has the source for all costs been described?
		Have discount rates been described and justified given the target decision maker?
D2d	Quality of life weights (utilities)	Are the utilities incorporated into the model appropriate?
		Is the source for the utility weights referenced?
		Are the methods of derivation for the utility weights justified?

D3	Data incorporation	Have all data incorporated into the model been described and referenced in sufficient detail?
		Has the use of mutually inconsistent data been justified?
		Is the process of data incorporation transparent?
		If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?
		If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?
D4	Assessment of uncertainty	Have the four principle types of uncertainty been justified?
		If not, has the omission of particular forms of uncertainty been justified?
D4a	Methodological	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?
D4b	Structural	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?
D4c	Heterogeneity	Has heterogeneity been dealt with by running the model separately for different subgroups?
D4d	Parameters	Are the methods of assessment of parameter uncertainty appropriated?
		If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?

Dimension of quality		**Questions for criteria appraisal**
C1	Internal consistency	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?
C2	External consistency	Are any counterintuitive results from the model explained and justified?
		If the model has been calibrated against independent data, have any differences been explained and justified?
		Have the results of the model been compared with those previous models and any differences in results explained?

Questions may be answered with “yes”, “no” or “non-applicable” Based on quality assessment (Philips et al., 2006)

Appendix IV: PIRSMA checklist of items to include when reporting a systematic review

Criteria of items to include when reporting a systematic review

Section/topic		Check list item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both
ABSTRACT		
Structured summary	2	Provide a structure summary including , as applicable; background; objective; data sources; study eligibility criteria, participants and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings, systematic review registration number.
INTRODUCTION		
Rational	3	Describe the rational for the review in the context of what is already known.
Objective	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)
METHODS		
Protocol and registration	5	Indicate of a review protocol exists, if and where it can be accessed (e.g., Web address), and , if available, provide registration information including registration number.
Eligibility sources	6	Specify study characteristic (e.g., PICOS, length of follow-up) and report characteristic (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (e.g., databases with date of coverage, contact with study authors to indentify additional studies), in the search and date last searched.
Search	8	Present full electronic search strategy for at least one databases, including any limits used, such that it could be reported.
Study selection	9	State the process for selection studies (i.g., screening, eligibility, included in systematic review, and if applicable, included in the meta-analysis).
Data collection and process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in supplicate) and any processes for obtaining and confirming data from investigators
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principle summary measures (e.g., risk ratio, difference in means)
Synthesis of the results	14	Describe the methods of handing data and combining results of studies, if done, including measures of consistency (e.g.,I2) for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publications bias, selective reporting with studies).
Additional analysis	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta –regression), if done, indicating which were pre-specified.
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and including in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristic for which data were extracted (e.g., study size, PICOS, follow-up, period) and provide the citations.
Risk of bias with in studies	19	Present data on risk of bias of each study and , if available, any outcome-level assessment (see item 12)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15).
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyse, meta-regression (see item 160).
DISCUSSION		
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).

Limitations	25	Discuss limitations at study and outcome level (e.g., incomplete retrieval of identified research, reporting bias).
	26	Provide a general interpretation of the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding for the systematic review and other support 9e.g., supply of data); role of funders for the systematic review.

Based on The PRISMA Statement for reporting systematic reviews and meta-analyses of studies the evaluated health care intervention: explanation and elaboration (Liberati et al., 2009)

Section/topic	#	Prevalence of GDM: Systematic review			Cost analysis and cost-effectiveness analysis: Systematic review		
		Sec	Page		Sec	Page	
TITLE							
Title	1	-	-	-	-	-	
ABSTRACT							
Structured summary	2	-	-	-	-	-	
INTRODUCTION							
Rational	3	5.1	106		6.1	124	
Objective	4	5.1	106	Para 3rd	6.1	124	Para 1st
METHODS							
Protocol and registration	5	-			-		
Eligibility sources	6	5.2.2	107		6.2.2	125	
Information sources	7	5.2.1	107		6.2.1	125	
Search	8	App I	272		App I	272	
Study selection	9	5.3 Figure 5.1	110		6.3	127	
Data collection and process	10	5.2.3	108		6.2.3	126	
Data items	11	Table 5.5	122		Table 6.4 Table 6.6	134, 141	
Risk of bias in individual studies	12	5.2.5	109		6.2.5	127	
Summary measures	13	5.2.5	109		6.2.5	127	
Synthesis of the results	14	5.2.5	109		6.2.5	127	
Risk of bias across studies	15	5.2.5	109		6.2.5	127	
Additional analysis	16	-			-		
RESULTS							
Study selection	17	5.3	110		6.3	127	
Study characteristics	18	5.5	111		6.5,6.6	131, 141	
Risk of bias with in studies	19	5.4	111		6.4	129	
Results of individual studies	20	Table 5.4, 5.5 Figure 5.2,5.3, 5.4	113,122, 114,115, 116		Table 6.4 Table 6.6	134, 141	
Synthesis of results	21	5.5	111		6.5,6.6	131, 135	
Risk of bias across studies	22	-			-		
Additional analysis	23	-			-		
DISCUSSION							
Summary of evidence	24	5.6	116		6.7	143	
Limitations	25	5.6.4	120		6.7.4	149	
Conclusions	26	5.7	121		6.8	149	
FUNDING							
Funding	27	-	-	-	-	-	-

Appendix V: Economic modelling of screening test for GDM (Chapter 6)

NICE (2008)

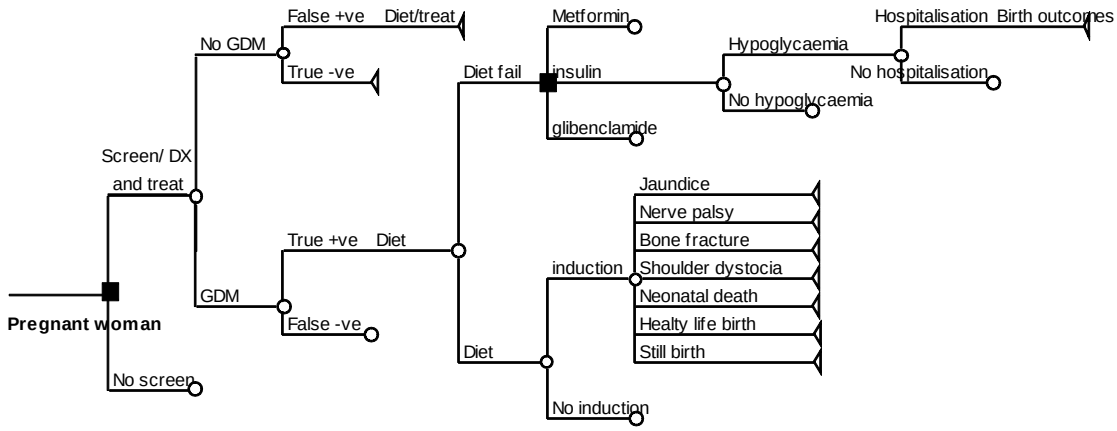
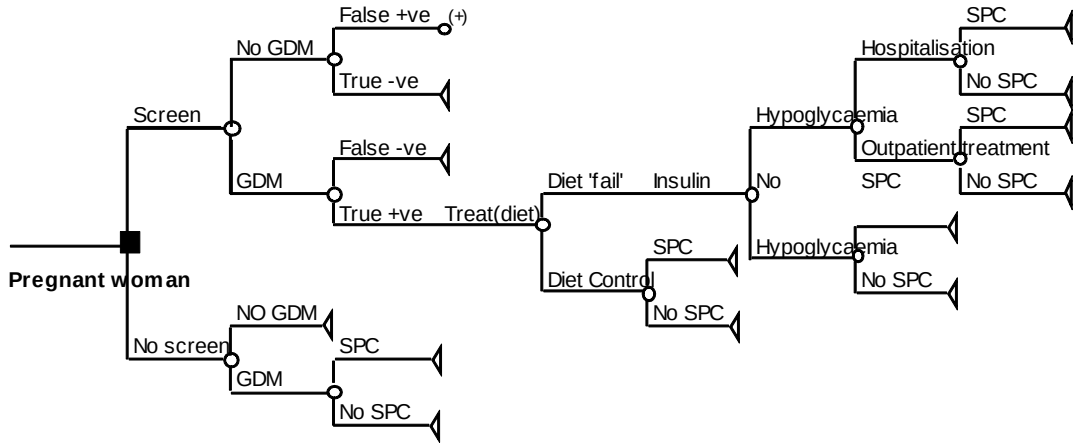


Figure 1 NICE basic decision tree model

Round et al., (2011)



SPC= Serious perinatal complication

[+] sub-tree callapsed but identical to treat-diet

Figure 2 Round et al basic decision tree model

Lohse et al., (2011)

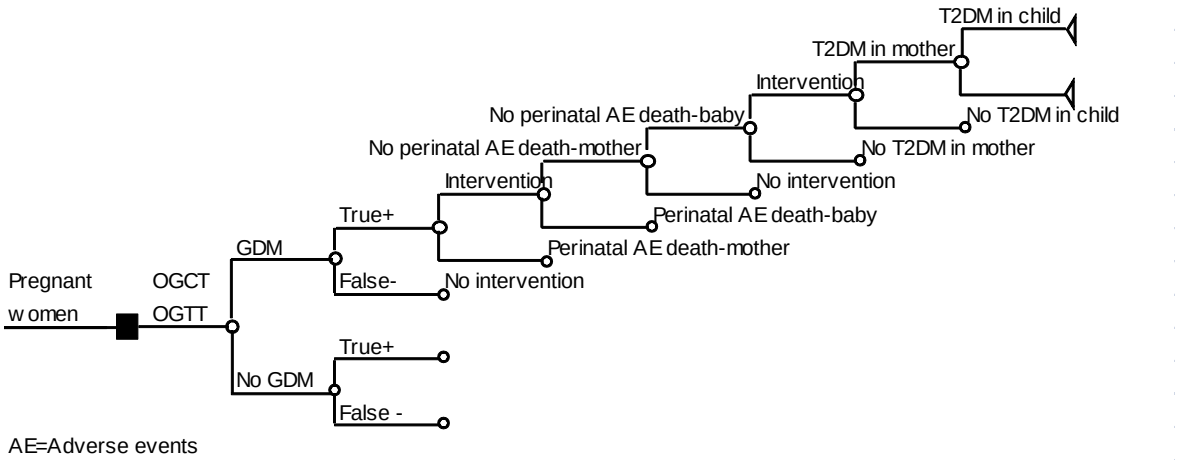


Figure 3 Lohse et al., basic decision tree model

Werner et al., (2012)

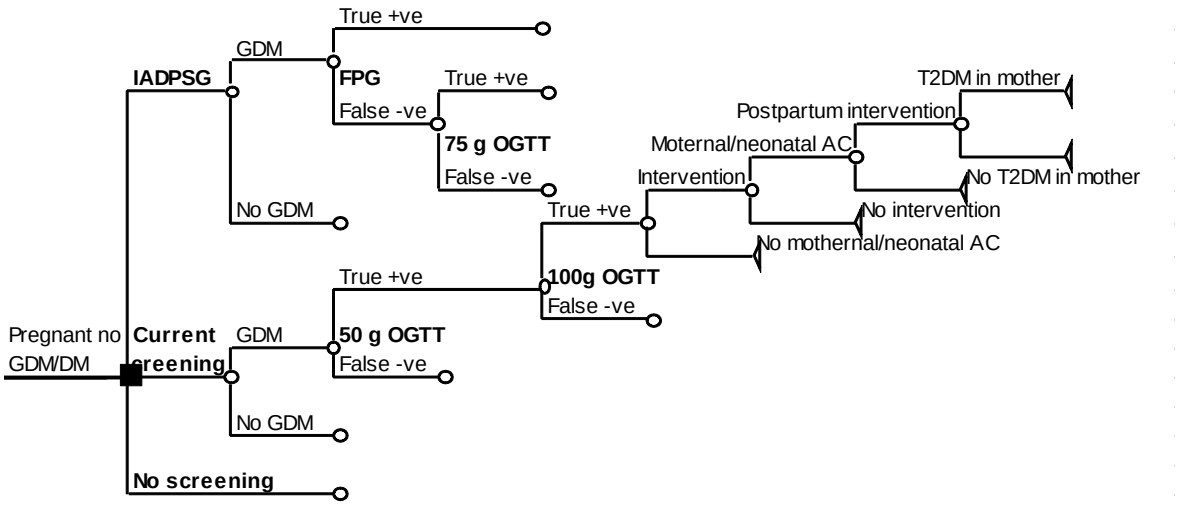


Figure 4 Werner et al., basic decision tree model

Mission et al.,(2012)

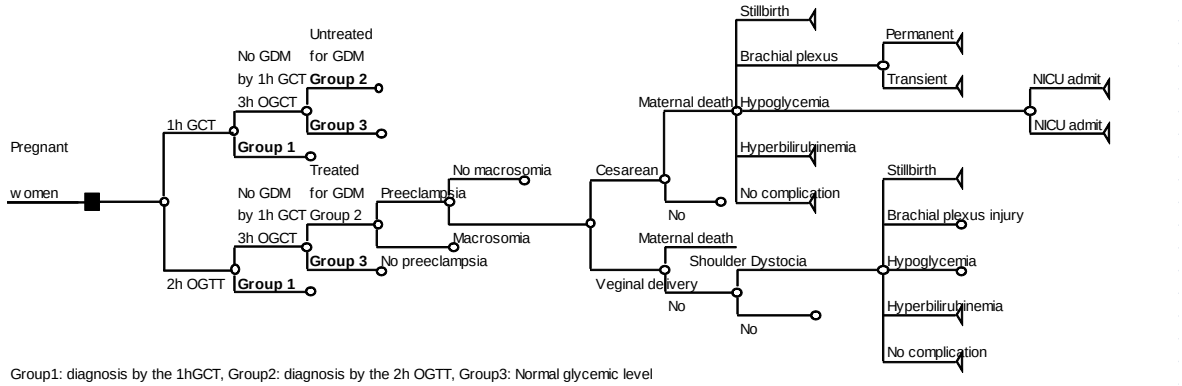


Figure 5 Mission et al., basic decision tree model

Marseille et al.,(2013)

This study was developed to assess the cost and health impact and cost-effectiveness of screening tests for GDM based on Lohse and colleagues 2011 model (Lohse et al., 2011). The same settings (India and Israel) and the decision tree model from Lohse's study were used, as showed in Figure 3.

Appendix VI: CHEERS checklist

Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist – Items to include when reporting economic evaluation of health interventions

Section/Item	Item no	Recommendation	Reported on page no. /line no
Title and abstract			
Title	1	Indentify the study as an economic evaluation, or use more specific terms such as "cost-effectiveness analysis and describe the intervention compared	
Abstract	2	Provide a structured summary of objective, perspective, setting, methods (including study design and inputs), results (including base-case and uncertainty analyses) and conclusions.	
Introduction			
Background and objective	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed including why they were chosen.	
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	
Study perspective	6	Describe the perspective of the study and related this to the cost being evaluated.	
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	
Time horizon	9	State the time horizons(s) over which costs and consequences are being evaluated and say why appropriate.	
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	
	11b	*Synthesis-based estimates:* Describe fully the methods used for the identification of included studies and synthesis of clinical effeteness data.	
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preference for outcomes.	
Estimating resources and costs	13a	*Single study-based evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resources item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe and adjustments made to approximate to opportunity costs.	
Currency, price date, and conversion	14	Report the dates of the estimated resources quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	
Choice of model	15	Describe and give reasons for the type of decision-analytic model used. Providing a figure to show model structure is strongly recommended.	
Assumptions	16	Describe all structure or other assumptions underpinning the decision-analytic model.	
Analytic methods	17	Describe all analytic methods supporting the evaluation, This could included methods for dealing with skewed, missing or censored data; extrapolation methods; methods for polling data; approaches to validate or make adjustments (e.g., half-cycle correction) to a model; and methods for handing population heterogeneity and uncertainty.	
Results			
Study parameters	18	Report the value, range reference, and if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values in strongly recommended.	
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated cost and outcomes of interest, as well as mean difference between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	
Characterizing uncertainty	20a	*Single study-based evaluation:* Describe the effects of sampling uncertainty for estimated incremental cost, incremental effectiveness, and incremental cost-effectiveness, together with the impact of	

			methodological assumptions (such as discount rate, study perspective)	
		20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions	
	Characterizing heterogeneity	21	If applicable, report differences in costs, outcome, or cost-effectiveness that can be explained by variations between subgroup of patients with different based line characteristics or other observed variability in effects that are not reducible by more information.	
Discussion				
	Study findings, Limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the finding and how the findings fit with current knowledge	
Other				
	Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other nonmonetary sources of support.	
	Conflicts of interest	24	Describe any potential for conflict of interest among study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors' recommendations.	

Section/Item	Item no	Case identification		Short term complications		Long term complicaitons	
		Reported on		Reported on		Reported on	
Title and abstract		Sec	Page	Sec	Page	Sec	Page
Title	1						
Abstract	2						
Introduction							
Background and objective	3	7.3	152	7.3	152	7.3	152
Methods							
Target population and subgroups	4	7.3.2	153	7.3.2	153	7.3.2	153
Setting and location	5						
Study perspective	6	7.3.1	153	7.3.1	153	7.3.1	153
Comparators	7	7.3.2	153	7.3.2	153	7.3.2	153
Time horizon	9	-		8.3	189	8.4	217
Discount rate	9	-		8.3	189	8.4	217
Choice of health outcomes	10			8.3.5	200	8.6	225
Measurement of effectiveness	11b	7.5.1	163	8.3.5	200	8.6	225
Measurement and valuation of preference based outcomes	12	-		8.3.5	200	8.6	225
Estimating resources and costs	13b	7.5.2	169	8.3.4	198	8.5	223
Currency, price date, and conversion	14	7.5.2	169	7.5.2	169	7.5.2	169
Choice of model	15	7.4	157	8.3.1, 8.3.2	190, 192	8.4.1, 8.4.2	218, 219
Assumptions	16	7.5.2	169	8.3.5	200	8.4.3, 8.7	220, 226
Analytic methods	17	7.6	171	8.3.8	213	8.7.3, 8.7.4	230, 231
Results							
Study parameters	18	7.5.1	136	8.3.3	193	8.4.3	220
Incremental costs and outcomes	19	7.7.3	178	9.2.1	233	9.2.2	236
Characterizing uncertainty	20b	7.7.4, 7.7.5	180, 180	9.3.1	240	9.3.2	240
Characterizing heterogeneity	21	-		-		-	
Discussion							
Study findings	22	7.8	182	9.4.1	241	9.4.2	242
Current knowledge	22	7.8	182	9.6.1	251	9.6.2	253
Limitations	22	10.4	265	10.4	265	10.4	265
Generalisability	22	10.5	266	10.5	266	10.5	266
Other							
Source of funding	23	-	-	-	-	-	-
Conflicts of interest	24	-	-	-	-	-	-

List of Reference

AARONSON, N. K., AHMEDZAI, S., BERGMAN, B., BULLINGER, M., CULL, A., DUEZ, N. J., FILIBERTI, A., FLECHTNER, H., FLEISHMAN, S. B., DE HAES, J. C. & ET AL. 1993. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst,* 85**,** 365-76.

ACOG 1994. *An update on Obstetrics and Gynecology*.

ACOG 2000. Practice Bulletin Clinical Management Guidelines:Fetal Macrosomia.

ACOG 2001. Practice Bulletin Clinical management guidelines for obstetrician-gynecologists. Number 30, September 2001 (replaces Technical Bulletin Number 200, December 1994). Gestational diabetes. *Obstet Gynecol,* 98**,** 525-38.

ADAMS, K. M., LI, H., NELSON, R. L., OGBURN, P. L., JR. & DANILENKO-DIXON, D. R. 1998. Sequelae of unrecognized gestational diabetes. *Am J Obstet Gynecol,* 178**,** 1321-32.

ADES, A. E., LU, G. & CLAXTON, K. 2004. Expected value of sample information calculations in medical decision modeling. *Med Decis Making,* 24**,** 207-27.

AGARWAL, M. M., DHATT, G. S. & OTHMAN, Y. 2012. Gestational diabetes in a tertiary care hospital: implications of applying the IADPSG criteria. *Arch Gynecol Obstet,* 286**,** 373-8.

AGARWAL, M. M., DHATT, G. S. & PUNNOSE, J. 2006. Gestational diabetes: an alternative, patient-friendly approach for using the diagnostic 100-g OGTT in high-risk populations. *Arch Gynecol Obstet,* 273**,** 325-30.

AGARWAL, M. M., HUGHES, P. F. & EZIMOKHAI, M. 2000. Screening for gestational diabetes in a high-risk population using fasting plasma glucose. *Int J Gynaecol Obstet,* 68**,** 147-8.

ALBERTI, K. G. & ZIMMET, P. Z. 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med,* 15**,** 539-53.

ALCOCER, L. & CUETO, L. 2008. Hypertension, a health economics perspective. *Ther Adv Cardiovasc Dis,* 2**,** 147-55.

ALTMAN, D. G. & BLAND, J. M. 1994a. Diagnostic tests 2: Predictive values. *BMJ,* 309**,** 102.

ALTMAN, D. G. & BLAND, J. M. 1994b. Diagnostic tests. 1: Sensitivity and specificity. *BMJ,* 308**,** 1552.

AMERICAN DIABETES, A. 2004. Gestational diabetes mellitus. *Diabetes Care,* 27 Suppl 1**,** S88-90.

AMERICAN DIABETES ASSOCIATION 1997. American Diabetes Association: clinical practice recommendations 1997. *Diabetes Care,* 20 Suppl 1**,** S1-70.

AMERICAN DIABETES ASSOCIATION 1998. Gestational diabetes mellitus. *Diabetes care,* 22 Suppl 21**,** 60-62.

AMERICAN DIABETES ASSOCIATION 2003. Gestational diabetes mellitus. *Diabetes Care,* 26 Suppl 1**,** S103-5.

AMERICAN DIABETES ASSOCIATION 2006. Diagnosis and classification of diabetes mellitus. *Diabetes Care,* 29 Suppl 1**,** S43-8.

ANDERSON, J. P., BUSH, J. W., CHEN, M. & DOLENC, D. 1986. Policy space areas and properties of benefit-cost/utility analysis. *JAMA,* 255**,** 794-5.

ANDREW, J. 2008. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. *AM Stat* 62**,** 314-320.

ANDRONIS, L., BARTON, P. & BRYAN, S. 2009. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technol Assess,* 13**,** iii, ix-xi, 1-61.

ANTHONY, K. 2007. Understanding diagnostic test 2: likelihood ratios, pre-and post-test probabilities and their use in clinical practice. *Acta Paediatrica* 96**,** 487-491.

ANZAKU, A. S. & MUSA, J. 2013. Prevalence and associated risk factors for gestational diabetes in Jos, North-central, Nigeria. *Arch Gynecol Obstet,* 287**,** 859-63.

BARDENHEIER, B. H., ELIXHAUSER, A., IMPERATORE, G., DEVLIN, H. M., KUKLINA, E. V., GEISS, L. S. & CORREA, A. 2013. Variation in prevalence of gestational diabetes mellitus among hospital discharges for obstetric delivery across 23 states in the United States. *Diabetes Care,* 36**,** 1209-14.

BARRY, C. D. & GABBE, S. 1998. Gestational Diabetes:Ditection, Management, and Implications. *Clinical diabetes* [Online], 16. Available: http://journal.diabetes.org/clinicaldiabetes/v16n1J-F98/pg4.htm [Accessed 30 Nov 2010].

BARTON, J. R., ISTWAN, N. B., RHEA, D., COLLINS, A. & STANZIANO, G. J. 2006. Cost-savings analysis of an outpatient management program for women with pregnancy-related hypertensive conditions. *Dis Manag,* 9**,** 236-41.

BARTON, P., BRYAN, S. & ROBINSON, S. 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy,* 9**,** 110-8.

BELLAMY, L., CASAS, J. P., HINGORANI, A. D. & WILLIAMS, D. 2009. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet,* 373**,** 1773-9.

BENTLEY-LEWIS, R., LEVKOFF, S., STUEBE, A. & SEELY, E. W. 2008. Gestational diabetes mellitus: postpartum opportunities for the diagnosis and prevention of type 2 diabetes mellitus. *Nat Clin Pract Endocrinol Metab,* 4**,** 552-8.

BERGER, H., CRANE, J., FARINE, D., ARMSON, A., DE LA RONDE, S., KEENAN-LINDSAY, L., LEDUC, L., REID, G. & VAN AERDE, J. 2002. Screening for gestational diabetes mellitus. *J Obstet Gynaecol Can,* 24**,** 894-912.

BILCKE, J., BEUTELS, P., BRISSON, M. & JIT, M. 2011. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. *Med Decis Making,* 31**,** 675-92.

BISHOP, F. L., PRESCOTT, P., CHAN, Y. K., SAVILLE, J., VON ELM, E. & LEWITH, G. T. 2010. Prevalence of complementary medicine use in pediatric cancer: a systematic review. *Pediatrics,* 125**,** 768-76.

BLACK, W. C. 1990. The CE plane: a graphic representation of cost-effectiveness. *Med Decis Making,* 10**,** 212-4.

BLEICHRODT, H. & JOHANNESSON, M. 1997. An experimental test of a theoretical foundation for rating-scale valuations. *Med Decis Making,* 17**,** 208-16.

BRAVO VERGEL, Y. & SCULPHER, M. 2008. Quality-adjusted life years. *Pract Neurol,* 8**,** 175-82.

BRAZIER, J., ROBERTS, J. & DEVERILL, M. 2002. The estimation of a preference-based measure of health from the SF-36. *J Health Econ,* 21**,** 271-92.

BRENNAN, A. & AKEHURST, R. 2000. Modelling in health economic evaluation. What is its place? What is its value? *Pharmacoeconomics,* 17**,** 445-59.

BRENNAN, A., CHICK, S. E. & DAVIES, R. 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ,* 15**,** 1295-310.

BRIGGS, A. 1998. *Uncertainty in the cost-effectiveness of health care interventions.* D.Phil, Oxford.

BRIGGS, A., CLAXTON, K. & SCULPHE, M. 2008. *Decision modelling for health economic evaluation*, Oxford University.

BRIGGS, A., CLAXTON, K. & SCULPHER, M. 2006. *Decision Modelling for Health Economic Evaluation*, Oxford university press.

BRIGGS, A. & FENN, P. 1998. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ,* 7**,** 723-40.

BRIGGS, A., SCULPHER, M. & BUXTON, M. 1994. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ,* 3**,** 95-104.

BRIGGS, A. H. 2000. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics,* 17**,** 479-500.

BRIGGS, A. H. 2007. Handing uncertainty in economic evaluaiton and presenting the results. *In:* DRUMMOND, M. F. & MCGUIRE, A. (eds.) *Economic evaluation in health care merging theory with practice.* New York: Oxford University.

BRIGGS, A. H. & GRAY, A. M. 1999. Handling uncertainty when performing economic evaluaiton of healthcare interventions. *Health techonogy assessement,* 3.

BRODY, S. C., HARRIS, R. & LOHR, K. 2003. Screening for gestational diabetes: a summary of the evidence for the U.S. Preventive Services Task Force. *Obstet Gynecol,* 101**,** 380-92.

BROWN, G. C., BROWN, M. M., SHARMA, S., BROWN, H., GOZUM, M. & DENTON, P. 2000. Quality of life associated with diabetes mellitus in an adult population. *J Diabetes Complications,* 14**,** 18-24.

BRYSON, C. L., IOANNOU, G. N., RULYAK, S. J. & CRITCHLOW, C. 2003. Association between gestational diabetes and pregnancy-induced hypertension. *Am J Epidemiol,* 158**,** 1148-53.

BUCHANAN, T. A., KJOS, S. L., MONTORO, M. N., WU, P. Y., MADRILEJO, N. G., GONZALEZ, M., NUNEZ, V., PANTOJA, P. M. & XIANG, A. 1994. Use of fetal ultrasound to select metabolic therapy for pregnancies complicated by mild gestational diabetes. *Diabetes Care,* 17**,** 275-83.

CADTH 2006. Guidelines for the Economic Evaluation of Health Technologies:Canada. 3 ed. Ottawa: Canadian Agency for Drugs and Technologies in Health.

CARPENTER, M. W. & COUSTAN, D. R. 1982. Criteria for screening tests for gestational diabetes. *Am J Obstet Gynecol,* 144**,** 768-73.

CHANPRAPAPH, P. & SUTJARIT, C. 2004. Prevalence of gestational diabetes mellitus (GDM) in women screened by glucose challenge test (GCT) at Maharaj Nakorn Chiang Mai Hospital. *J Med Assoc Thai,* 87**,** 1141-6.

CHAUHAN, S. P., GROBMAN, W. A., GHERMAN, R. A., CHAUHAN, V. B., CHANG, G., MAGANN, E. F. & HENDRIX, N. W. 2005. Suspicion and treatment of the macrosomic fetus: a review. *Am J Obstet Gynecol,* 193**,** 332-46.

CHEN, Y., QUICK, W. W., YANG, W., ZHANG, Y., BALDWIN, A., MORAN, J., MOORE, V., SAHAI, N. & DALL, T. M. 2009. Cost of gestational diabetes mellitus in the United States in 2007. *Popul Health Manag,* 12**,** 165-74.

CIANNI D G, FATATI. F, LAPOLLA. A, LEOTTA. S, MANNINO. D, PARILLO. M & PIPICELLI. G 2008. Dietary therapy in diabetec pregnancy: recommendations. *Mediterr Jnutr Metab,* 1**,** 49-60.

CLARKE, P., GRAY, A., LEGOOD, R., BRIGGS, A. & HOLMAN, R. 2003. The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65). *Diabet Med,* 20**,** 442-50.

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. & GROUP, U. K. P. D. S. 2004. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia,* 47**,** 1747-59.

CLAUSEN, T. D., MATHIESEN, E. R., HANSEN, T., PEDERSEN, O., JENSEN, D. M., LAUENBORG, J. & DAMM, P. 2008. High prevalence of type 2 diabetes and pre-diabetes in adult offspring of women with gestational diabetes mellitus or type 1 diabetes: the role of intrauterine hyperglycemia. *Diabetes Care,* 31**,** 340-6.

CLAXTON, K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ,* 18**,** 341-64.

CLAXTON, K., EGGINGTON, S., GINNELLY, L., GRIFFIN, S., MCCABE, C., PHILIPS, Z., TAPPENDEN, P. & WAILOO, A. 2005a. A Pilot stusy of value of informaion analysis to support research recommendations for the national institute for clinical excellence. Available: http://www.york.ac.uk/che/pdf/claxtonnice.pdf [Accessed 05 May].

CLAXTON, K., GINNELLY, L., SCULPHER, M., PHILIPS, Z. & PALMER, S. 2004. A pilot study on the use of decision theory and value of information analysis as part

of the NHS Health Technology Assessment programme. *Health Technol Assess,* 8**,** 1-103, iii.

CLAXTON, K., SCULPHER, M. & DRUMMOND, M. 2002. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet,* 360**,** 711-5.

CLAXTON, K., SCULPHER, M., MCCABE, C., BRIGGS, A., AKEHURST, R., BUXTON, M., BRAZIER, J. & O'HAGAN, T. 2005b. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ,* 14**,** 339-47.

CLEMENTS, R. V. 2001. *Shoulder dystocia: Risk management and litigation in Obstetrics and Gynaecology,* London, RSM Press in association with RCOG Press.

CLINICAL RESOURCE EFFICIENCY SUPPORT TEAM 2001. Management of diabetes in pregnancy. Belfast.

COUSINS, L., DATTEL, B. J., HOLLINGSWORTH, D. R. & ZETTNER, A. 1984. Glycosylated hemoglobin as a screening test for carbohydrate intolerance in pregnancy. *Am J Obstet Gynecol,* 150**,** 455-60.

COUSTAN, D. R. 1991. Screening and diagnosis of gestational diabetes. *Isr J Med Sci,* 27**,** 503-9.

COUSTAN, D. R. & CARPENTER, M. W. 1998. The diagnosis of gestational diabetes. *Diabetes Care,* 21 Suppl 2**,** B5-8.

COUSTAN, D. R., CARPENTER, M. W., O'SULLIVAN, P. S. & CARR, S. R. 1993. Gestational diabetes: predictors of subsequent disordered glucose metabolism. *Am J Obstet Gynecol,* 168**,** 1139-44; discussion 1144-5.

COUSTAN, D. R. & LEWIS, S. B. 1978. Insulin therapy for gestational diabetes. *Obstet Gynecol,* 51**,** 306-10.

CROWTHER, C. A., HILLER, J. E., MOSS, J. R., MCPHEE, A. J., JEFFRIES, W. S. & ROBINSON, J. S. 2005. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med,* 352**,** 2477-86.

CUNNINGHAM, S. J. 2001. An introduction to economic evaluation of health care. *J Orthod,* 28**,** 246-50.

DE VECIANA, M., MAJOR, C. A., MORGAN, M. A., ASRAT, T., TOOHEY, J. S., LIEN, J. M. & EVANS, A. T. 1995. Postprandial versus preprandial blood glucose monitoring in women with gestational diabetes mellitus requiring insulin therapy. *N Engl J Med,* 333**,** 1237-41.

DEPARTMENT OF HEALTH 2010. Spending Review 2010. UK: The Stationery Office Limited.

DI CIANNI, G., VOLPE, L., CASADIDIO, I., BOTTONE, P., MARSELLI, L., LENCIONI, C., BOLDRINI, A., TETI, G., DEL PRATO, S. & BENZI, L. 2002. Universal screening and intensive metabolic management of gestational diabetes: cost-effectiveness in Italy. *Acta Diabetol,* 39**,** 69-73.

DIABETES CONTROL AND COMPLICATIONS TRIAL RESEARCH GROUP 1996. Influence of intensive diabetes treatment on quality-of-life outcomes in the diabetes control and complications trial. *Diabetes Care,* 19**,** 195-203.

DIABETES PREVENTION AND CONTROL PROGRAM. 2007. Gestational Diabetes Nutrition Guidelines. Available: http://www.colorado.gov/cs/Satellite?blobcol=urldata&blobheadername1=Content-Disposition&blobheadername2=Content-Type&blobheadervalue1=inline%3B+filename%3D%22Colorado+Nutrition+Guidelines+on+Gestational+Diabetes.pdf%22&blobheadervalue2=application%2Fpdf&blobkey=id&blobtable=MungoBlobs&blobwhere=1251811774389&ssbinary=true [Accessed 1 Oct 2013].

DIABETES UK. 2010. Diabetes in The UK 2010: Key statistics on diabetes. Available: http://www.diabetes.org.uk/Documents/Reports/Diabetes_in_the_UK_2010.pdf [Accessed 11 March 2011].

DIABETES UK. 2012. Diabetes in The UK 2012: Key statistics on diabetes. Available: http://www.diabetes.org.uk/Documents/Reports/Diabetes-in-the-UK-2012.pdf [Accessed 11 July 2013].

DORNHORST, A., PATERSON, C. M., NICHOLLS, J. S., WADSWORTH, J., CHIU, D. C., ELKELES, R. S., JOHNSTON, D. G. & BEARD, R. W. 1992. High prevalence of gestational diabetes in women from ethnic minority groups. *Diabet Med,* 9**,** 820-5.

DRUMMOND, M. F. & JEFFERSON, T. O. 1996. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ,* 313**,** 275-83.

DRUMMOND, M. F. & MCGUIRE, A. 2007. *Economic evaluation in healht care,* Oxford, Oxford press.

DRUMMOND, M. F., SCULPHER, M.J., TORRANCE, G. W. O'BRIEN, B. AND STODDART, G.L. 1987. *Methods for the economic evaluation of health care programmes,* Oxford, Oxford University Press.

DRUMMOND, M. F., SCULPHER, M.J., TORRANCE, G. W. O'BRIEN, B. AND STODDART, G.L. 2005a. *Methods for economic evaluation of health care programes,* Oxford, Oxford University Press.

DRUMMOND, M. F., SCULPHER, M.J., TORRANCE, G. W. O'BRIEN, B. AND STODDART, G.L. 2005b. *Methods for the economic evaluation of health care programmes,* Oxford, Oxford University Press.

DRUMMOND, M. F., STODDART, G. L. & TORRANCE, G. W. 1987 *Methods for the economic evaluation of health care programmes,* Oxford, Oxford University Press.

ECKERMANN, S., KARNON, J. & WILLAN, A. R. 2010. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics,* 28**,** 699-709.

EDDY, D. M. 1990. Clinical decision making: from theory to practice. Designing a practice policy. Standards, guidelines, and options. *JAMA,* 263**,** 3077, 3081, 3084.

EDGAR., R. B., DONALD., R. B., THOMAS., G. T. & ROBERT., J. P. 1999. *Diagnostic Strategies for Common Medical Problems,* Pennsylvania, American College of Physicians-Amarican Society of Internal Medicine.

ELLIOTT, B. D., SCHENKER, S., LANGER, O., JOHNSON, R. & PRIHODA, T. 1994. Comparative placental transport of oral hypoglycemic agents in humans: a model of human placental drug transfer. *Am J Obstet Gynecol,* 171**,** 653-60.

ENGELGAU, M. M., HERMAN, W. H., SMITH, P. J., GERMAN, R. R. & AUBERT, R. E. 1995. The epidemiology of diabetes and pregnancy in the U.S., 1988. *Diabetes Care,* 18**,** 1029-33.

EVANS, C., TAVAKOLI, M. & CRAWFORD, B. 2004. Use of quality adjusted life years and life years gained as benchmarks in economic evaluations: a critical appraisal. *Health Care Manag Sci,* 7**,** 43-9.

FEIG, D. S., ZINMAN, B., WANG, X. & HUX, J. E. 2008. Risk of development of diabetes mellitus after diagnosis of gestational diabetes. *CMAJ,* 179**,** 229-34.

FENWICK, E. & BYFORD, S. 2005. A guide to cost-effectiveness acceptability curves. *Br J Psychiatry,* 187**,** 106-8.

FENWICK, E., CLAXTON, K. & SCULPHER, M. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ,* 10**,** 779-87.

FENWICK, E., MARSHALL, D. A., LEVY, A. R. & NICHOL, G. 2006. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *BMC Health Serv Res,* 6**,** 52.

FERRARA, A., HEDDERSON, M. M., QUESENBERRY, C. P. & SELBY, J. V. 2002. Prevalence of gestational diabetes mellitus detected by the national diabetes data group or the carpenter and coustan plasma glucose thresholds. *Diabetes Care,* 25**,** 1625-30.

FERRARA, A. & KIM, C. 2009. Gestational Diabetes Mellitus:Diagnosis Matenal and Fetal Outcomes and Management. *In:* TASTSOULTS, A., WYCKOFF, J. A. & FLORENCE, M. B. (eds.) *Diabetes in Women:Pathophysiology and Therapy.* New York: Springer Science+Bussiness Media.

FERRARA, A., WEISS, N. S., HEDDERSON, M. M., QUESENBERRY, C. P., JR., SELBY, J. V., ERGAS, I. J., PENG, T., ESCOBAR, G. J., PETTITT, D. J. & SACKS, D. A. 2007. Pregnancy plasma glucose levels exceeding the American Diabetes Association thresholds, but below the National Diabetes Data Group thresholds for gestational diabetes mellitus, are related to the risk of neonatal macrosomia, hypoglycaemia and hyperbilirubinaemia. *Diabetologia,* 50**,** 298-306.

FIRTH, G. R. 1996. *Digbetes and Pregnancy An International Approach to Diagnosis and Management,* London, John Willey & Sons.

FORGER, F., OSTENSEN, M., SCHUMACHER, A. & VILLIGER, P. M. 2005. Impact of pregnancy on health related quality of life evaluated prospectively in pregnant women with rheumatic diseases by the SF-36 health survey. *Ann Rheum Dis,* 64**,** 1494-9.

FOX, C. & PICKERING, A. 1995. *Diabetis in the real world,* London, Clays

GARDNER, I. A., STRYHN, H., LIND, P. & COLLINS, M. T. 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev Vet Med,* 45**,** 107-22.

GARNER, P., OKUN, N., KEELY, E., WELLS, G., PERKINS, S., SYLVAIN, J. & BELCHER, J. 1997. A randomized controlled trial of strict glycemic control and tertiary level obstetric care versus routine obstetric care in the management of gestational diabetes: a pilot study. *Am J Obstet Gynecol,* 177**,** 190-5.

GENERAL REGISTER OFFICE FOR SCOTLAND. 2011. Life Expectancy in Scottish Parliamentary Constituency Areas, 2008-2010.

GETAHUN, D., NATH, C., ANANTH, C. V., CHAVEZ, M. R. & SMULIAN, J. C. 2008. Gestational diabetes in the United States: temporal trends 1989 through 2004. *Am J Obstet Gynecol,* 198**,** 525 e1-5.

GHERMAN, R. B., OUZOUNIAN, J. G., MILLER, D. A., KWOK, L. & GOODWIN, T. M. 1998. Spontaneous vaginal delivery: a risk factor for Erb's palsy? *Am J Obstet Gynecol,* 178**,** 423-7.

GILBERT, W. M., NESBITT, T. S. & DANIELSEN, B. 2003. The cost of prematurity: quantification by gestational age and birth weight. *Obstet Gynecol,* 102**,** 488-92.

GILLESPIE, P., O'NEILL, C., AVALOS, G., DUNNE, F. P. & COLLABORATORS, A. D. 2012. New estimates of the costs of universal screening for gestational diabetes mellitus in Ireland. *Ir Med J,* 105**,** 15-8.

GILLESPIE, P., O'NEILL, C., AVALOS, G., O'REILLY, M., DUNNE, F. & COLLABORATORS, A. D. 2011. The cost of universal screening for gestational diabetes mellitus in Ireland. *Diabet Med,* 28**,** 912-8.

GILLMAN, M. W., OAKEY, H., BAGHURST, P. A., VOLKMER, R. E., ROBINSON, J. S. & CROWTHER, C. A. 2010. Effect of treatment of gestational diabetes mellitus on obesity in the next generation. *Diabetes Care,* 33**,** 964-8.

GLASZIOU, P. P. & SANDERS, S. L. 2002. Investigating causes of heterogeneity in systematic reviews. *Stat Med,* 21**,** 1503-11.

GODWIN, M., MUIRHEAD, M., HUYNH, J., HELT, B. & GRIMMER, J. 1999. Prevalence of gestational diabetes mellitus among Swampy Cree women in Moose Factory, James Bay. *CMAJ,* 160**,** 1299-302.

GOLD, M. R., SIEGEL, J. E., RUSSELL, L. B. & WEINSTEIN, M. C. 1996. *Cost-effeciveness in health and medicine,* New York, Oxford University Press.

GOLD, M. R., STEVENSON, D. & FRYBACK, D. G. 2002. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annu Rev Public Health,* 23**,** 115-34.

GRECO, P., LOVERRO, G. & SELVAGGI, L. 1994. Does gestational diabetes represent an obstetrical risk factor? *Gynecol Obstet Invest,* 37**,** 242-5.

GREENE, M. F. 1997. Screening for gestational diabetes mellitus. *N Engl J Med,* 337**,** 1625-6.

GREENHALGH, T. 1997. How to read a paper. Papers that report diagnostic or screening tests. *BMJ,* 315**,** 540-3.

HADAEGH, F., TOHIDI, M., HARATI, H., KHEIRANDISH, M. & RAHIMI, S. 2005. Prevalence of gestational diabetes mellitus in southern Iran (Bandar Abbas City). *Endocr Pract,* 11**,** 313-8.

HALKOAHO, A., KAVILO, M., PIETILA, A. M., HUOPIO, H., SINTONEN, H. & HEINONEN, S. 2010. Does gestational diabetes affect women's health-related quality of life after delivery? *Eur J Obstet Gynecol Reprod Biol,* 148**,** 40-3.

HANNA, F. W. & PETERS, J. R. 2002. Screening for gestational diabetes; past, present and future. *Diabet Med,* 19**,** 351-8.

HANNA, F. W., PETERS, J. R., HARLOW, J. & JONES, P. W. 2008. Gestational diabetes screening and glycaemic management; national survey on behalf of the Association of British Clinical Diabetologists. *QJM,* 101**,** 777-84.

HANNINEN, J., TAKALA, J. & KEINANEN-KIUKAANNIEMI, S. 1998. Quality of life in NIDDM patients assessed with the SF-20 questionnaire. *Diabetes Res Clin Pract,* 42**,** 17-27.

HAPO 2009. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with neonatal anthropometrics. *Diabetes,* 58**,** 453-9.

HARDEO, S. 1992. Teaching Bayes's Theorem using Examples in Medical Diagnosis. *Teaching mathematics and its applications,* 11**,** 175-179.

HAY, J. & JACKSON, J. 1999. Panel 2: methodological issues in conducting pharmacoeconomic evaluations--modeling studies. *Value Health,* 2**,** 78-81.

HELTON, M. R., ARNDT, J., KEBEDE, M. & KING, M. 1997. Do low-risk prenatal patients really need a screening glucose challenge test? *J Fam Pract,* 44**,** 556-61.

HERMIDA, R. C., AYALA, D. E., MOJON, A., FERNANDEZ, J. R., ALONSO, I., SILVA, I., UCIEDA, R. & IGLESIAS, M. 2000. Blood pressure patterns in normal pregnancy, gestational hypertension, and preeclampsia. *Hypertension,* 36**,** 149-58.

HIDEN, U. & DESOYE, G. 2010. Insulin and the placenta in GDM. *In:* KIM, C. & FERRARA, A. (eds.) *Gestational diabetes during and after pregnancy.* London: Springer.

HINE, D. 1999. For the good that it will do: issues confronting healthcare in the UK. *J R Soc Med,* 92**,** 332-8.

HOCH, J. S., BRIGGS, A. H. & WILLAN, A. R. 2002. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ,* 11**,** 415-30.

HOEDJES, M. 2011. *Maternal Qualtiy of Life,Lifestyle, and Interventions after complicated Pregnancies.* Master, University Rotterdam.

HOFFMAN, L. 1998. Gestational diabetes mellitus (GDM). *Med J Aust,* 168**,** 140.

HOFFMAN, L., NOLAN, C., WILSON, J. D., OATS, J. J. & SIMMONS, D. 1998. Gestational diabetes mellitus--management guidelines. The Australasian Diabetes in Pregnancy Society. *Med J Aust,* 169**,** 93-7.

HORVATH, K., KOCH, K., JEITLER, K., MATYAS, E., BENDER, R., BASTIAN, H., LANGE, S. & SIEBENHOFER, A. 2010. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. *BMJ,* 340**,** c1395.

HOSSEIN-NEZHAD, A., MAGHBOOLI, Z., VASSIGH, A. R. & LARIJANI, B. 2007. Prevalence of gestational diabetes mellitus and pregnancy outcomes in Iranian women. *Taiwan J Obstet Gynecol,* 46**,** 236-41.

HOSSEING-NEZHAD, A., MAGHBOOL, Z., VASSIGH, A. R. & LARIJANI, B. 2007. Prevalence of gestational diabetes mellitus and pregnancy outcomes in Iranian women. *Taiwai J Obstet Gynecol,* 46**,** 236-241.

HUNKING, M., GLAZIOU, P., SIEGEL, J., WEEKS, J., PLISKIN, J. & ELSTEIN, A. 2001. *Decision making in health and medicine. Integrating evidence and values,* Cambridge, Cambridge University Press.

HUSEREAU, D., DRUMMOND, M., PETROU, S., CARSWELL, C., MOHER, D., GREENBERG, D., AUGUSTOVSKI, F., BRIGGS, A. H., MAUSKOPF, J., LODER, E. & FORCE, I. H. E. E. P. G.-C. G. R. P. T. 2013. Consolidated Health

Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health,* 16**,** 231-50.

JANGHORBANI, M., STENHOUSE, E., JONES, R. B. & MILLWARD, A. 2006. Gestational diabetes mellitus in Plymouth, U.K.: prevalence, seasonal variation and associated factors. *J Reprod Med,* 51**,** 128-34.

JARRETT, R. J., CASTRO-SOARES, J., DORNHORST, A. & BEARD, R. W. 1997. Should we screen for gestational diabetes? *BMJ,* 315**,** 736-9.

JENSEN, D. M., SORENSEN, B., FEILBERG-JORGENSEN, N., WESTERGAARD, J. G. & BECK-NIELSEN, H. 2000. Maternal and perinatal outcomes in 143 Danish women with gestational diabetes mellitus and 143 controls with a similar risk profile. *Diabet Med,* 17**,** 281-6.

JIMENEZ-MOLEON, J. J., BUENO-CAVANILLAS, A., LUNA-DEL-CASTILLO, J. D., GARCIA-MARTIN, M., LARDELLI-CLARET, P. & GALVEZ-VARGAS, R. 2002. Prevalence of gestational diabetes mellitus: variations related to screening strategy used. *Eur J Endocrinol,* 146**,** 831-7.

JIWANI, A., MARSEILLE, E., LOHSE, N., DAMM, P., HOD, M. & KAHN, J. G. 2012. Gestational diabetes mellitus: results from a survey of country prevalence and practices. *J Matern Fetal Neonatal Med,* 25**,** 600-10.

JOHANNESSON, M., JONSSON, B. & KARLSSON, G. 1996. Outcome measurement in economic evaluation. *Health Econ,* 5**,** 279-96.

JONES, C. W. 2001. Gestational diabetes and its impact on the neonate. *Neonatal Netw,* 20**,** 17-23.

JOVANOVIC, L. 1998. American Diabetes Association's Fourth International Workshop-Conference on Gestational Diabetes Mellitus: summary and discussion. Therapeutic interventions. *Diabetes Care,* 21 Suppl 2**,** B131-7.

JOVANOVIC, L. 2010. Screening and diagnosis of diabetes melliuts during pregnancy. Available: http://www.uptodate.com/patients/content/topic.do?topicKey=~V0MSg9jLKsJx3jl [Accessed 27/10/10].

JYOTI, E. & RICHARD, H. 2009. Screening and diagnostic tests. Available: http://emedicine.medscape.come/aritcle/773832 [Accessed 28/04/10].

KARCAALTINCABA, D., BUYUKKARAGOZ, B., KANDEMIR, O., YALVAC, S., KIYKAC-ALTINBAS, S. & HABERAL, A. 2011. Gestational diabetes and gestational impaired glucose tolerance in 1653 teenage pregnancies: prevalence, risk factors and pregnancy outcomes. *J Pediatr Adolesc Gynecol,* 24**,** 62-5.

KARCAALTINCABA, D., KANDEMIR, O., YALVAC, S., GUVENDAG-GUVEN, S. & HABERAL, A. 2009. Prevalence of gestational diabetes mellitus and gestational impaired glucose tolerance in pregnant women evaluated by National Diabetes Data Group and Carpenter and Coustan criteria. *Int J Gynaecol Obstet,* 106**,** 246-9.

KARLSSON, G. & JOHANNESSON, M. 1996. The decision rules of cost-effectiveness analysis. *Pharmacoeconomics,* 9**,** 113-20.

KARNON, J. 2003. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ,* 12**,** 837-48.

KARNON, J. & BROWN, J. 1998. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manag Sci,* 1**,** 133-40.

KENNEDY, L., IDRIS, I. & GAZIS, A. 2006. Ploblem Solving in Diabetes. Oxford: Clinical publishing.

KHAMIS, H. 1987. Statistics refresher: test of hypothesis and diagnostic test evaluation *JDMS,* 3**,** 123-129.

KHAMIS, H. J. 1990. An application of Bayes' Rule to Diagnostic test evaluation. *JDMS,* 4**,** 212-218.

KIM, C., BERGER, D. K. & CHAMANY, S. 2007a. Recurrence of gestational diabetes mellitus: a systematic review. *Diabetes Care,* 30**,** 1314-9.

KIM, C., HERMAN, W. H. & VIJAN, S. 2007b. Efficacy and cost of postpartum screening strategies for diabetes among women with histories of gestational diabetes mellitus. *Diabetes Care,* 30**,** 1102-6.

KIM, C., NEWTON, K. M. & KNOPP, R. H. 2002. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care,* 25**,** 1862-8.

KIM, S. Y., GOLDIE, S. J. & SALOMON, J. A. 2010. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. *Med Decis Making,* 30**,** E1-E28.

KING, H. 1998. Epidemiology of glucose intolerance and gestational diabetes in women of childbearing age. *Diabetes Care,* 21 Suppl 2**,** B9-13.

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.

KING, P. 1996. *The EuroQol instrument: an index of health-related quality of life. In Spilker B(ed). Qualtiy of life and Pharmacoeconomics in Clinical Trials,* Philadelphia, Lippincott-Raven.

KITZMILLER, J. L. 2010. An overview of problems and solutions in the diagnosis and treatment fo gestational diabetes. *In:* KIM, C. & FERRARA, A. (eds.) *Gestational Diabetes During and After Pregnancy.* London: Springer.

KITZMILLER, J. L., DANG-KILDUFF, L. & TASLIMI, M. M. 2007. Gestational diabetes after delivery. Short-term management and long-term risks. *Diabetes Care,* 30 Suppl 2**,** S225-35.

KITZMILLER, J. L., ELIXHAUSER, A., CARR, S., MAJOR, C. A., DE VECIANA, M., DANG-KILDUFF, L. & WESCHLER, J. M. 1998. Assessment of costs and benefits of management of gestational diabetes mellitus. *Diabetes Care,* 21 Suppl 2**,** B123-30.

KJOS, S. L. & BUCHANAN, T. A. 1999. Gestational diabetes mellitus. *N Engl J Med,* 341**,** 1749-56.

KJOS, S. L., BUCHANAN, T. A., GREENSPOON, J. S., MONTORO, M., BERNSTEIN, G. S. & MESTMAN, J. H. 1990. Gestational diabetes mellitus: the prevalence of glucose intolerance and diabetes mellitus in the first two months post partum. *Am J Obstet Gynecol,* 163**,** 93-8.

KOUKKOU, E., TAUB, N., JACKSON, P., METCALFE, G., CAMERON, M. & LOWY, C. 1995. Difference in prevalence of gestational diabetes and perinatal outcome in an innercity multiethnic London population. *Eur J Obstet Gynecol Reprod Biol,* 59**,** 153-7.

LAKING, G., LORD, J. & FISCHER, A. 2006. The economics of diagnosis. *Health Econ,* 15**,** 1109-20.

LANDON, M. B., SPONG, C. Y., THOM, E., CARPENTER, M. W., RAMIN, S. M., CASEY, B., WAPNER, R. J., VARNER, M. W., ROUSE, D. J., THORP, J. M., JR., SCISCIONE, A., CATALANO, P., HARPER, M., SAADE, G., LAIN, K. Y., SOROKIN, Y., PEACEMAN, A. M., TOLOSA, J. E. & ANDERSON, G. B. 2009. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med,* 361**,** 1339-48.

LANGER, O. 1993. Management of gestational diabetes. *Clin Perinatol,* 20**,** 603-17.

LANGER, O., CONWAY, D. L., BERKUS, M. D., XENAKIS, E. M. & GONZALES, O. 2000. A comparison of glyburide and insulin in women with gestational diabetes mellitus. *N Engl J Med,* 343**,** 1134-8.

LANGER, O., YOGEV, Y., XENAKIS, E. M. & ROSENN, B. 2005. Insulin and glyburide therapy: dosage, severity level of gestational diabetes, and pregnancy outcome. *Am J Obstet Gynecol,* 192**,** 134-9.

LARIJANI, B., HOSSEIN-NEZHAD, A., RIZVI, S. W., MUNIR, S. & VASSIGH, A. R. 2003. Cost analysis of different screening strategies for gestational diabetes mellitus. *Endocr Pract,* 9**,** 504-9.

LE LAY, A., DESPIEGEL, N., FRANCOIS, C. & DURU, G. 2006. Can discrete event simulation be of use in modelling major depression? *Cost Eff Resour Alloc,* 4**,** 19.

LEMEN, P. M., WIGTON, T. R., MILLER-MCCARTHEY, A. J. & CRUIKSHANK, D. P. 1998. Screening for gestational diabetes mellitus in adolescent pregnancies. *Am J Obstet Gynecol,* 178**,** 1251-6.

LESAFFRE, E., SPEYBROECK, N. & BERKVENS, D. 2007. Bayes and diagnostic testing. *Vet Parasitol,* 148**,** 58-61.

LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GOTZSCHE, P. C., IOANNIDIS, J. P., CLARKE, M., DEVEREAUX, P. J., KLEIJNEN, J. & MOHER, D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol,* 62**,** e1-34.

LIM-UY, S. W., CUNANAN, E. C. & ANDAG-SILVA, A. A. 2010. Prevalence and Risk Factors of Gestational Diabetes Mellitus at the

University of Santo Tomas Hospital. *Philippine Journal of Internal Medicine,* 48 24-31.

LIND, T. & PHILLIPS, P. R. 1991. Influence of pregnancy on the 75-g OGTT. A prospective multicenter study. The Diabetic Pregnancy Study Group of the European Association for the Study of Diabetes. *Diabetes,* 40 Suppl 2**,** 8-13.

LOHSE, N., MARSEILLE, E. & KAHN, J. G. 2011. Development of a model to assess the cost-effectiveness of gestational diabetes mellitus screening and lifestyle change for the prevention of type 2 diabetes mellitus. *Int J Gynaecol Obstet,* 115 Suppl 1**,** S20-5.

LORGELLY, P. K., LAWSON, K. D., FENWICK, E. A. & BRIGGS, A. H. 2010. Outcome measurement in economic evaluations of public health interventions: a role for the capability approach? *Int J Environ Res Public Health,* 7**,** 2274-89.

LOUKINE, L., WATERS, C., CHOI, B. C. & ELLISON, J. 2012. Impact of diabetes mellitus on life expectancy and health-adjusted life expectancy in Canada. *Popul Health Metr,* 10**,** 7.

MAGEE, M. S., KNOPP, R. H. & BENEDETTI, T. J. 1990. Metabolic effects of 1200-kcal diet in obese pregnant women with gestational diabetes. *Diabetes,* 39**,** 234-40.

MAGEE, M. S., WALDEN, C. E., BENEDETTI, T. J. & KNOPP, R. H. 1993. Influence of diagnostic criteria on the incidence of gestational diabetes and perinatal morbidity. *JAMA,* 269**,** 609-15.

MAJOR, C. A., DEVECIANA, M., WEEKS, J. & MORGAN, M. A. 1998. Recurrence of gestational diabetes: who is at risk? *Am J Obstet Gynecol,* 179**,** 1038-42.

MARCIA, A. T. & DONALD, C. S. 1996. Assessment of Quality-of-Life Outcomes. *N Engl J Med* 334**,** 835-840.

MARQUETTE, G. P., KLEIN, V. R. & NIEBYL, J. R. 1985. Efficacy of screening for gestational diabetes. *Am J Perinatol,* 2**,** 7-9.

MARSEILLE, E., LOHSE, N., JIWANI, A., HOD, M., SESHIAH, V., YAJNIK, C. S., ARORA, G. P., BALAJI, V., HENRIKSEN, O., LIEBERMAN, N., CHEN, R., DAMM, P., METZGER, B. E. & KAHN, J. G. 2013. The cost-effectiveness of gestational diabetes screening including prevention of type 2 diabetes: application of a new model in India and Israel. *J Matern Fetal Neonatal Med,* 26**,** 802-10.

MARTINE, H., MARIEKE, K. & ANJOKE, J. 2007. Gestational diabetes : A Review of the current literature and guidelines. *Obstetrical and Gynecological Survey,* 62**,** 125-136.

MATZA, L. S., BOYE, K. S., YURGIN, N., BREWSTER-JORDAN, J., MANNIX, S., SHORR, J. M. & BARBER, B. L. 2007. Utilities and disutilities for type 2 diabetes treatment-related attributes. *Qual Life Res,* 16**,** 1251-65.

MCFARLANE, P. A. & BAYOUMI, A. M. 2004. Acceptance and rejection: cost-effectiveness and the working nephrologist. *Kidney Int,* 66**,** 1735-41.

MCMASTER UNIVERSITY HEALTH AND SCIENCES CENTRE 1981. How to read clinical journals: II. To learn about a diagnostic test. *Can Med Assoc J,* 124**,** 703-10.

MEDICINENET. 2004. *Definition of Bayes theorem* [Online]. New York City: MedicineNet.com Available: http://www.medterms.com/script/main/art.asp?articlekey=10301.

MELTZER, S. J., SNYDER, J., PENROD, J. R., NUDI, M. & MORIN, L. 2010. Gestational diabetes mellitus screening and diagnosis: a prospective randomised controlled trial comparing costs of one-step and two-step methods. *BJOG,* 117**,** 407-15.

METZGER, B. E. 1991. Summary and recommendations of the Third International Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes,* 40 Suppl 2**,** 197-201.

METZGER, B. E., BUCHANAN, T. A., COUSTAN, D. R., DE LEIVA, A., DUNGER, D. B., HADDEN, D. R., HOD, M., KITZMILLER, J. L., KJOS, S. L., OATS, J. N., PETTITT, D. J., SACKS, D. A. & ZOUPAS, C. 2007. Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes Care,* 30 Suppl 2**,** S251-60.

METZGER, B. E. & COUSTAN, D. R. 1998. Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee. *Diabetes Care,* 21 Suppl 2**,** B161-7.

METZGER, B. E., GABBE, S. G., PERSSON, B., BUCHANAN, T. A., CATALANO, P. A., DAMM, P., DYER, A. R., LEIVA, A., HOD, M., KITZMILER, J. L., LOWE, L. P., MCINTYRE, H. D., OATS, J. J., OMORI, Y. & SCHMIDT, M. I. 2010a. Internation Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy. *Diabet Care,* 33**,** 676-682.

METZGER, B. E., GABBE, S. G., PERSSON, B., BUCHANAN, T. A., CATALANO, P. A., DAMM, P., DYER, A. R., LEIVA, A., HOD, M., KITZMILER, J. L., LOWE, L. P., MCINTYRE, H. D., OATS, J. J., OMORI, Y. & SCHMIDT, M. I. 2010b. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care,* 33**,** 676-82.

MILNE, F., REDMAN, C., WALKER, J., BAKER, P., BRADLEY, J., COOPER, C., DE SWIET, M., FLETCHER, G., JOKINEN, M., MURPHY, D., NELSON-PIERCY, C., OSGOOD, V., ROBSON, S., SHENNAN, A., TUFFNELL, A., TWADDLE, S. & WAUGH, J. 2005. The pre-eclampsia community guideline (PRECOG): how to screen for and detect onset of pre-eclampsia in the community. *BMJ,* 330**,** 576-80.

MIRES, G. J., WILLIAMS, F. L. & HARPER, V. 1999. Screening practices for gestational diabetes mellitus in UK obstetric units. *Diabet Med,* 16**,** 138-41.

MISSION, J. F., OHNO, M. S., CHENG, Y. W. & CAUGHEY, A. B. 2012. Gestational diabetes screening with the new IADPSG guidelines: a cost-effectiveness analysis. *Am J Obstet Gynecol,* 207**,** 326 e1-9.

MOORE, T. R. 2007. Glyburide for the treatment of gestational diabetes. A critical appraisal. *Diabetes Care,* 30 Suppl 2**,** S209-13.

MOSES, R., FULWOOD, S. & GRIFFITHS, R. 1997. Gestational diabetes mellitus; resource utilization and costs of diagnosis and treatment. *Aust N Z J Obstet Gynaecol,* 37**,** 184-6.

MOSES, R., GRIFFITHS, R. & DAVIS, W. 1995. Gestational diabetes: do all women need to be tested? *Aust N Z J Obstet Gynaecol,* 35**,** 387-9.

MOSES, R. G. 2010. New consensus criteria for GDM: problem solved or a pandora's box? *Diabetes Care,* 33**,** 690-1.

MOSES, R. G., MORRIS, G. J., PETOCZ, P., SAN GIL, F. & GARG, D. 2011. The impact of potential new diagnostic criteria on the prevalence of gestational diabetes mellitus in Australia. *Med J Aust,* 194**,** 338-40.

MOSS, J. R., CROWTHER, C. A., HILLER, J. E., WILLSON, K. J., ROBINSON, J. S. & AUSTRALIAN CARBOHYDRATE INTOLERANCE STUDY IN PREGNANT WOMEN, G. 2007. Costs and consequences of treatment for mild gestational

diabetes mellitus - evaluation from the ACHOIS randomised trial. *BMC Pregnancy Childbirth,* 7**,** 27.
MOYNIHAN, R., DOUST, J. & HENRY, D. 2012. Preventing overdiagnosis: how to stop harming the healthy. *BMJ,* 344**,** e3502.
MUSHLIN, A. I., RUCHLIN, H. S. & CALLAHAN, M. A. 2001. Costeffectiveness of diagnostic tests. *Lancet,* 358**,** 1353-5.
NATIONAL DIABETES DATA GROUP 1979. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National Diabetes Data Group. *Diabetes,* 28**,** 1039-57.
NATIONAL HEALTH SERVICE. 1948. *NHA core principles* [Online]. Available: http://www.nhs.uk/NHSEngland/thenhs/about/Pages/nhscoreprinciples.aspx.
NATIONAL HEALTH SERVICE 2008a. Antenatal care diabetes in pregnancy costing report implementing NICE guidance. London.
NATIONAL HEALTH SERVICE 2008b. Antenatal care Diabetes in pregnancy Costing report Implementing NICE guidance. London: National Institute for Health and Clinical Excellence
NATIONAL INSTITUTE OF DIABETES AND DIGESTIVE AND KIDNEY DISEASES 2008. National Diabetes Statistics, 2007 fact sheet. Bethesda, MD: U.S. Department of Health and Human Services.
NAYLOR, C. D., SERMER, M., CHEN, E. & FARINE, D. 1997. Selective screening for gestational diabetes mellitus. Toronto Trihospital Gestational Diabetes Project Investigators. *N Engl J Med,* 337**,** 1591-6.
NCSSM STATISTICS LEADERSHIP INSTITUTE. 1999. Sensitivity and Specificity. *Statistics leadership Institute.Categorical Data Analysis* [Online]. Available: http://www.ncssm.edu/courses/math/Stat_Inst/Stats2007/Stat%20and%20Calc/Sensitivity%20and%20Specificity.pdf.
NHS 2013. Scottish Perinatal and Infant Mortality and Morbidity Report 2011. Edinburgh: NHSScotland.
NHS EED. 2007. *Idenfifying studies for inclusion in NHS EED* [Online]. York: University of York. Available: http://www.york.ac.uk/inst/crd/intertasc/nhs_eed_strategies.html [Accessed 12 August 2012].
NHS SCOTLAND 2012. Scottish Diabetes Survey 2011. Edinburgh: Scottish Diabetes Survey Monitoring Group.
NHS SCOTLAND. 2013. *Scottish Medicines Consortium* [Online]. Available: https://www.scottishmedicines.org.uk/About_SMC/Policy_Statements/Policy_Statements [Accessed 26 August 2014].
NHS TARIFF 06/07 2007. Department of Health: National Tarriff 2006/7. 2006 ed.
NHS TARIFF 10/11 2011. NHS trust reference cost schedules 2010-11. 17 November 2011 ed.: NHS.
NHS. 2011. Pre-Existing Diabetes in pregnancy. *In:* WAI., L. (ed.).
NICE 2005. A guide to NICE. London: Abba Litho.
NICE 2008a. Costing template for Antenatal care, incorporation Diabetes in pregnancy. London: National Institute for Health and Clinical Excellece.
NICE. 2008b. Diabetes in pregnancy: Management of diabetes and its complications from pre-conception to the postnatal period. Available: http://www.ncbi.nlm.nih.gov/books/NBK51920/pdf/TOC.pdf [Accessed 29 Dec 2012].
NICE. 2008c. Guide to the methods of technology appraisal 2008. Available: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [Accessed 27 Feb 2010].
NICE 2008d. *NICE Clinical guideline 63: Diabetes in pregnancy Management of diabetes and its complications from pre-conception to the postnatal period,* London.
NICE 2011. Hypertension in pregnancy: The management of hypertensive disorders during pregnancy, NICE clinical guideline 107. National Institute for Health and Clinical Excellence

NICE. 2012. *Methods for the development of NICE public health guidance (third edition)* [Online]. Available: http://publications.nice.org.uk/methods-for-the-development-of-nice-public-health-guidance-third-edition-pmg4 [Accessed 26 Jun 2014].

NICE. 2013. Guide to the methods of technology appraisal 2013. NICE. Available: http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf [Accessed 22 July 2013].

NICHOLSON, W. K., FLEISHER, L. A., FOX, H. E. & POWE, N. R. 2005. Screening for gestational diabetes mellitus: a decision and cost-effectiveness analysis of four screening strategies. *Diabetes Care,* 28**,** 1482-4.

NOVIELLI, N., COOPER, N. J., ABRAMS, K. R. & SUTTON, A. J. 2010. How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997. *Value Health,* 13**,** 952-7.

NUIJTEN, M. J., PRONK, M. H., BRORENS, M. J., HEKSTER, Y. A., LOCKEFEER, J. H., DE SMET, P. A., BONSEL, G. & VAN DER KUY, A. 1998. Reporting format for economic evaluation. Part II: Focus on modelling studies. *Pharmacoeconomics,* 14**,** 259-68.

O'BRIEN, B. J., ANDERSON, D. R. & GOEREE, R. 1994a. Cost-effectiveness of enoxaparin versus warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *CMAJ,* 150**,** 1083-90.

O'BRIEN, B. J., DRUMMOND, M. F., LABELLE, R. J. & WILLAN, A. 1994b. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care,* 32**,** 150-63.

O'SULLIVAN, J. B. & MAHAN, C. M. 1964. Criteria for the Oral Glucose Tolerance Test in Pregnancy. *Diabetes,* 13**,** 278-85.

OFFICE FOR NATIONAL STATISTICS 2012. Ethnicity and National identity in England and Wales 2011. London: Office for national statistics.

OFFICE FOR NATIONAL STATISTICS 2013. Interim Life Tables, England and Waless,2009-2011. 21 March 2013 ed.

OKEH, U. M. & UGWU, A. C. 2008. Bayes' theorem: A paradigm research tool in biomedical science. *African Journal of Biotechnology,* 7**,** 4807-4818.

OOSTENBRINK, J. B., RUTTEN-VAN MOLKEN, M. P., MONZ, B. U. & FITZGERALD, J. M. 2005. Probabilistic Markov model to assess the cost-effectiveness of bronchodilator therapy in COPD patients in different countries. *Value Health,* 8**,** 32-46.

PALMER, S. & SMITH, P. C. 2000. Incorporating option values into the economic evaluation of health care technologies. *J Health Econ,* 19**,** 755-66.

PARMIGIANI, G. 2002. *Modeling in Medical Decision Making : A Bayesian Approach,* England, TJ International.

PAUKER, S. G. & KASSIRER, J. P. 1975. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med,* 293**,** 229-34.

PAUL, K., HARDMAN, G. & MACRAN, S. 1999. UK Populations Norms for EQ-5D. York: Center for Health Economics.

PEDERSEN, J. 1954. Weight and length at birth of infants of diabetic mothers. *Acta Endocrinol (Copenh),* 16**,** 330-42.

PEDERSEN, M. L., JACOBSEN, J. L. & JORGENSEN, M. E. 2010. Prevalence of gestational diabetes mellitus among women born in Greenland: measuring the effectiveness of the current screening procedure. *Int J Circumpolar Health,* 69**,** 352-60.

PERSSON, B. & HANSON, U. 1998. Neonatal morbidities in gestational diabetes mellitus. *Diabetes Care,* 21 Suppl 2**,** B79-84.

PERUCCHINI, D., FISCHER, U., SPINAS, G. A., HUCH, R., HUCH, A. & LEHMANN, R. 1999. Using fasting plasma glucose concentrations to screen for gestational diabetes mellitus: prospective population based study. *BMJ,* 319**,** 812-5.

PETER., G. C. 2007. *Rational Diagnosis and Treatment Evidence- Based Clinical Decision-Making,* Copenhagen, John Wiley & Sons, Ltd.

PETTITT, D. J. 2001. The 75-g oral glucose tolerance test in pregnancy. *Diabetes Care,* 24**,** 1129.

PETTITT, D. J., KNOWLER, W. C., BAIRD, H. R. & BENNETT, P. H. 1980. Gestational diabetes: infant and maternal complications of pregnancy in relation to third-trimester glucose tolerance in the Pima Indians. *Diabetes Care,* 3**,** 458-64.

PHELPS, C. E. & MUSHLIN, A. I. 1988. Focusing technology assessment using medical decision theory. *Med Decis Making,* 8**,** 279-89.

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 assessment. *Pharmacoeconomics,* 24**,** 355-71.

PHILIPS, Z., GINNELLY, L., SCULPHER, M., CLAXTON, K., GOLDER, S., RIEMSMA, R., WOOLACOOT, N. & GLANVILLE, J. 2004. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess,* 8**,** iii-iv, ix-xi, 1-158.

PHILLIPS, C. 2009. What is cost-effectiveness?

PHILLIPS, C. & THOMPSON, G. 2009. What is a QALY? Available: www.whatisseries.co.uk.

PONCET, B., TOUZET, S., ROCHER, L., BERLAND, M., ORGIAZZI, J. & COLIN, C. 2002. Cost-effectiveness analysis of gestational diabetes mellitus screening in France. *Eur J Obstet Gynecol Reprod Biol,* 103**,** 122-9.

POYHONEN-ALHO, M. K., TERAMO, K. A., KAAJA, R. J. & HIILESMAA, V. K. 2005. 50gram oral glucose challenge test combined with risk factor-based screening for gestational diabetes. *Eur J Obstet Gynecol Reprod Biol,* 121**,** 34-7.

PRINS, J., BLANKER, M. H., BOHNEN, A. M., THOMAS, S. & BOSCH, J. L. 2002. Prevalence of erectile dysfunction: a systematic review of population-based studies. *Int J Impot Res,* 14**,** 422-32.

PSSRU 2010. *Unit Costs of Health and Social Care 2010,* Kent, Personal Social Services Research Unit.

PUBLIC HEALTH ACTION SUPPORT TEAM , D. O. H. 2010. *Differences between screening and diagnostic, case finding* [Online]. London. Available: http://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnosis [Accessed 05/10 2010].

REDEKOP, W. K., KOOPMANSCHAP, M. A., STOLK, R. P., RUTTEN, G. E., WOLFFENBUTTEL, B. H. & NIESSEN, L. W. 2002. Health-related quality of life and treatment satisfaction in Dutch patients with type 2 diabetes. *Diabetes Care,* 25**,** 458-63.

REICHELT, A. J., SPICHLER, E. R., BRANCHTEIN, L., NUCCI, L. B., FRANCO, L. J. & SCHMIDT, M. I. 1998. Fasting plasma glucose is a useful test for the detection of gestational diabetes. Brazilian Study of Gestational Diabetes (EBDG) Working Group. *Diabetes Care,* 21**,** 1246-9.

REY, E. 1999. Screening for gestational diabetes mellitus. A simple test may make it easier to study whether screening is worthwhile. *BMJ,* 319**,** 798-9.

RICART, W., LOPEZ, J., MOZAS, J., PERICOT, A., SANCHO, M. A., GONZALEZ, N., BALSELLS, M., LUNA, R., CORTAZAR, A., NAVARRO, P., RAMIREZ, O., FLANDEZ, B., PALLARDO, L. F., HERNANDEZ, A., AMPUDIA, J., FERNANDEZ-REAL, J. M. & CORCOY, R. 2005. Potential impact of American Diabetes Association (2000) criteria for diagnosis of gestational diabetes mellitus in Spain. *Diabetologia,* 48**,** 1135-41.

RICHARDSON, G. & MANCA, A. 2004. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Econ,* 13**,** 1203-10.

RICHARDSON, W. S. & DETSKY, A. S. 1995. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA,* 273**,** 1292-5.

RISKIN-MASHIAH, S., DAMTI, A., YOUNES, G. & AUSLENDER, R. 2010. First trimester fasting hyperglycemia as a predictor for the development of gestational diabetes mellitus. *Eur J Obstet Gynecol Reprod Biol,* 152**,** 163-7.

ROBERTS, M., RUSSELL, L. B., PALTIEL, A. D., CHAMBERS, M., MCEWAN, P., KRAHN, M. & FORCE, I.-S. M. G. R. P. T. 2012a. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making,* 32**,** 678-89.

ROBERTS, T. E., BARTON, P. M., AUGUSTE, P. E., MIDDLETON, L. J., FURMSTON, A. T. & EWER, A. K. 2012b. Pulse oximetry as a screening test for congenital heart defects in newborn infants: a cost-effectiveness analysis. *Arch Dis Child,* 97**,** 221-6.

RODRIGUES, S., ROBINSON, E. & GRAY-DONALD, K. 1999. Prevalence of gestational diabetes mellitus among James Bay Cree women in northern Quebec. *CMAJ,* 160**,** 1293-7.

ROPER, N. A., BILOUS, R. W., KELLY, W. F., UNWIN, N. C., CONNOLLY, V. M. & SOUTH TEES DIABETES MORTALITY, S. 2002. Cause-specific mortality in a population with diabetes: South Tees Diabetes Mortality Study. *Diabetes Care,* 25**,** 43-8.

ROUND, J. A., JACKLIN, P., FRASER, R. B., HUGHES, R. G., MUGGLESTONE, M. A. & HOLT, R. I. 2011. Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease. *Diabetologia,* 54**,** 256-63.

ROWAN, J. A., GAO, W., HAGUE, W. M. & MCINTYRE, H. D. 2010. Glycemia and its relationship to outcomes in the metformin in gestational diabetes trial. *Diabetes Care,* 33**,** 9-16.

ROWAN, J. A., HAGUE, W. M., GAO, W., BATTIN, M. R. & MOORE, M. P. 2008. Metformin versus insulin for the treatment of gestational diabetes. *N Engl J Med,* 358**,** 2003-15.

ROWEN, D., BRAZIER, J. & ROBERTS, J. 2009. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes,* 7**,** 27.

RUF M. & MORGAN O. 2008. *Difference between screening and -diagnostic tests, case finding* [Online]. Available: http://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnostic/2c-diagnosis-screening/screening-diagnostic-case-finding [Accessed 4 April 2010].

RUSSELL, L. B., GOLD, M. R., SIEGEL, J. E., DANIELS, N. & WEINSTEIN, M. C. 1996. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA,* 276**,** 1172-7.

RYAN, E. A. 2011. Diagnosing gestational diabetes. *Diabetologia,* 54**,** 480-6.

SACKS, D. A., ABU-FADIL, S., KARTEN, G. J., FORSYTHE, A. B. & HACKETT, J. R. 1987. Screening for gestational diabetes with the one-hour 50-g glucose test. *Obstet Gynecol,* 70**,** 89-93.

SANGHERA, S., ORLANDO, R. & ROBERTS, T. 2013. Economic evaluations and diagnostic testing: an illustrative case study approach. *Int J Technol Assess Health Care,* 29**,** 53-60.

SANTINI, D. L. & ALES, K. L. 1990. The impact of universal screening for gestational glucose intolerance on outcome of pregnancy. *Surg Gynecol Obstet,* 170**,** 427-36.

SATTLEY, M. 2008. *The history of diabetes* [Online]. Available: http://diabeteshealth.com/read/2008/12/17/715/the-history-of-diabetes/?isComment=1 [Accessed 24 Sep 2013].

SAVONA-VENTURA, C. & CHIRCOP, M. 2003. The threshold criteria for the 75g oral glucose tolerance test in pregnancy and short-term adverse pregnancy outcomes. *Int J Diabetes & Metabolism,* 11.

SCHMIDT, M. I., DUNCAN, B. B., REICHELT, A. J., BRANCHTEIN, L., MATOS, M. C., COSTA E FORTI, A., SPICHLER, E. R., POUSADA, J. M., TEIXEIRA, M. M. & YAMASHITA, T. 2001. Gestational diabetes mellitus diagnosed with a 2-h 75-g

oral glucose tolerance test and adverse pregnancy outcomes. *Diabetes Care,* 24**,** 1151-5.
SCHMIDT, M. I., MATOS, M. C., REICHELT, A. J., FORTI, A. C., DE LIMA, L. & DUNCAN, B. B. 2000. Prevalence of gestational diabetes mellitus--do the new WHO criteria make a difference? Brazilian Gestational Diabetes Study Group. *Diabet Med,* 17**,** 376-80.
SCOTT, D. A., LOVEMAN, E., MCINTYRE, L. & WAUGH, N. 2002. Screening for gestational diabetes: a systematic review and economic evaluation. *Health Technol Assess,* 6**,** 1-161.
SCULPHER, M., DRUMMOND, M. & BUXTON, M. 1997. The iterative use of economic evaluation as part of the process of health technology assessment. *J Health Serv Res Policy,* 2**,** 26-30.
SCULPHER, M., FENWICK, E. & CLAXTON, K. 2000. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics,* 17**,** 461-77.
SENANAYAKE, H., SENEVIRATNE, S., ARIYARATNE, H. & WIJERATNE, S. 2006. Screening for gestational diabetes mellitus in southern Asian women. *J Obstet Gynaecol Res,* 32**,** 286-91.
SERLIN, D. C. & LASH, R. W. 2009. Diagnosis and management of gestational diabetes mellitus. *Am Fam Physician,* 80**,** 57-62.
SERMER, M., NAYLOR, C. D., GARE, D. J., KENSHOLE, A. B., RITCHIE, J. W., FARINE, D., COHEN, H. R., MCARTHUR, K., HOLZAPFEL, S., BIRINGER, A. & ET AL. 1995. Impact of increasing carbohydrate intolerance on maternal-fetal outcomes in 3637 women without gestational diabetes. The Toronto Tri-Hospital Gestational Diabetes Project. *Am J Obstet Gynecol,* 173**,** 146-56.
SEYOUM, B., KIROS, K., HAILESELASE, T. & LEOLE, A. 1999. Prevalence of gestational diabetes mellitus in rural pregnant mothers in northern Ethiopia. *Diabetes Res Clin Pract,* 46**,** 247-51.
SHEPARD, D. S. & THOMPSON, M. S. 1979. First principles of cost-effectiveness analysis in health. *Public Health Rep,* 94**,** 535-43.
SHIH, Y. C. & HALPERN, M. T. 2008. Economic evaluations of medical care interventions for cancer patients: how, why, and what does it mean? *CA Cancer J Clin,* 58**,** 231-44.
SIBAI, B., DEKKER, G. & KUPFERMINC, M. 2005. Pre-eclampsia. *Lancet,* 365**,** 785-99.
SIEBERT, U. 2003. When should decision-analytic modeling be used in the economic evaluation of health care? *Eur J Health Econom,* 4**,** 143-150.
SIGN 2001. *Management of diabetes A natioanl clinical guideline,* Edingurgh.
SIGN 2010. *Management of diabetes A national clinical guideline,* Edinburgh.
SIGN. 2002. Guide 57: Cardiac Rehabilitation A nationl clinical guideline. Edinburgh.
ŠIMUNDIĆ, M. 2008. Measure of diagnostic accuracy: basic definitions. *Medical and Biological Sciences,* 22**,** 61-65.
SINGH, B., HAWTHORNE, G. & VOS, T. 2001. The role of economic evaluation in mental Health care. *Australian and New Zealand Journal of Psychiatry***,** 104-117.
SOARES J DE AC, D. A., BEARD RW 1997. The case for screening for gestational diabetes. ***BMJ****,* 315**,** 737-739.
SOLLI, O., STAVEM, K. & KRISTIANSEN, I. S. 2010. Health-related quality of life in diabetes: The associations of complications with EQ-5D scores. *Health Qual Life Outcomes,* 8**,** 18.
SOONTHORNPUN, S., SOONTHORNPUN, K., AKSONTEING, J. & THAMPRASIT, A. 2003. A comparison between a 75-g and 100-g oral glucose tolerance test in pregnant women. *Int J Gynaecol Obstet,* 81**,** 169-73.
STINNETT, A. A. & PALTIEL, A. D. 1997. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. *Med Decis Making,* 17**,** 483-9.
SUMEKSRI, P., WONGYAI, S. & AIMPUN, P. 2006. Prevalence of gestational diabetes mellitus (GDM) in pregnant women aged 30 to 34 years old at Phramongkutklao Hospital. *J Med Assoc Thai,* 89 Suppl 4**,** S94-9.

SYMON, A. 2003. A review of mothers' prenatal and postnatal quality of life. *Health Qual Life Outcomes,* 1**,** 38.

TAN, J. M., MACARIO, A., CARVALHO, B., DRUZIN, M. L. & EL-SAYED, Y. Y. 2010. Cost-effectiveness of external cephalic version for term breech presentation. *BMC Pregnancy Childbirth,* 10**,** 3.

TANG, M. L. 2004. On simultaneous assessment of sensitivity and specificity when combining two diagnostic tests. *Stat Med,* 23**,** 3593-605.

TAPPENDEN, P., CHILCOTT, J. B., EGGINGTON, S., OAKLEY, J. & MCCABE, C. 2004. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess,* 8**,** iii, 1-78.

TAYLOR, M. 2009. What is sensitivity analysis ? Available: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_sens_analy.pdf.

TAYLOR, R. S., DRUMMOND, M. F., SALKELD, G. & SULLIVAN, S. D. 2004. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. *BMJ,* 329**,** 972-5.

TESTA, M. A., SIMONSON, D. C. & TURNER, R. R. 1998. Valuing quality of life and improvements in glycemic control in people with type 2 diabetes. *Diabetes Care,* 21 Suppl 3**,** C44-52.

THE SCOTTISH GOVERNMENT. 2011. *Ethnic Group Demographics* [Online]. The scottish government. Available: http://www.scotland.gov.uk/Topics/People/Equality/Equalities/DataGrid/Ethnicity/EthPopMig [Accessed 17 May 2014 2014].

THOMPSON, D. J., PORTER, K. B., GUNNELLS, D. J., WAGNER, P. C. & SPINNATO, J. A. 1990. Prophylactic insulin in the management of gestational diabetes. *Obstet Gynecol,* 75**,** 960-4.

TORKAN, B., PARSAY, S., LAMYIAN, M., KAZEMNEJAD, A. & MONTAZERI, A. 2009. Postnatal quality of life in women after normal vaginal delivery and caesarean section. *BMC Pregnancy Childbirth,* 9**,** 4.

TORRANCE, G. W., THOMAS, W. H. & SACKETT, D. L. 1972. A utility maximization model for evaluation of health care programs. *Health Serv Res,* 7**,** 118-33.

TREVISOL, D. J., MOREIRA, L. B., KERKHOFF, A., FUCHS, S. C. & FUCHS, F. D. 2011. Health-related quality of life and hypertension: a systematic review and meta-analysis of observational studies. *J Hypertens,* 29**,** 179-88.

TUROK, D. K., RATCLIFFE, S. D. & BAXLEY, E. G. 2003. Management of gestational diabetes mellitus. *Am Fam Physician,* 68**,** 1767-72.

VALDERRABANO, F., JOFRE, R. & LOPEZ-GOMEZ, J. M. 2001. Quality of life in end-stage renal disease patients. *Am J Kidney Dis,* 38**,** 443-64.

VAN LEEUWEN, M., OPMEER, B. C., YILMAZ, Y., LIMPENS, J., SERLIE, M. J. & MOL, B. W. 2010. Accuracy of the random glucose test as screening test for gestational diabetes mellitus: a systematic review. *Eur J Obstet Gynecol Reprod Biol,* 154**,** 130-5.

VIS, A. N., KRANSE, R., ROOBOL, M., VAN DER KWAST, T. H. & SCHRODER, F. H. 2002. Serendipity in detecting disease in low prostate-specific antigen ranges. *BJU Int,* 89**,** 384-9.

VOGEL, N., BURNAND, B., VIAL, Y., RUIZ, J., PACCAUD, F. & HOHLFELD, P. 2000. Screening for gestational diabetes: variation in guidelines. *Eur J Obstet Gynecol Reprod Biol,* 91**,** 29-36.

VOS, T., CORRY, J., HABY, M. M., CARTER, R. & ANDREWS, G. 2005. Cost-effectiveness of cognitive-behavioural therapy and drug interventions for major depression. *Aust N Z J Psychiatry,* 39**,** 683-92.

WAGNER, L. K. 2004. Diagnosis and management of preeclampsia. *Am Fam Physician,* 70**,** 2317-24.

WARE, J. E., JR. & SHERBOURNE, C. D. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care,* 30**,** 473-83.

WARE, J. E., SNOW, K. K., KOSINSKI, M. & GANDEK, B. 1993. *SF-36 health survey manual and interpretation guide,* Boston, New England Medical Center, The Health Institute.

WARNER, J. 2004. Clinicians's guide to evaluating diagnostic and screening tests in psychiatry. *Advance in Psyschiatric Treatment* 10**,** 446-454.

WEEKS, J. W., MAJOR, C. A., DE VECIANA, M. & MORGAN, M. A. 1994. Gestational diabetes: does the presence of risk factors influence perinatal outcome? *Am J Obstet Gynecol,* 171**,** 1003-7.

WEINSTEIN, M. C., O'BRIEN, B., HORNBERGER, J., JACKSON, J., JOHANNESSON, M., MCCABE, C. & LUCE, B. R. 2003. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health,* 6**,** 9-17.

WEINTROB, N., KARP, M. & HOD, M. 1996. Short- and long-range complications in offspring of diabetic mothers. *J Diabetes Complications,* 10**,** 294-301.

WERNER, E. F., HAN, C. S., PETTKER, C. M., BUHIMSCHI, C. S., COPEL, J. A., FUNAI, E. F. & THUNG, S. F. 2011. Universal cervical-length screening to prevent preterm birth: a cost-effectiveness analysis. *Ultrasound Obstet Gynecol,* 38**,** 32-7.

WERNER, E. F., PETTKER, C. M., ZUCKERWISE, L., REEL, M., FUNAI, E. F., HENDERSON, J. & THUNG, S. F. 2012. Screening for gestational diabetes mellitus: are the criteria proposed by the international association of the Diabetes and Pregnancy Study Groups cost-effective? *Diabetes Care,* 35**,** 529-35.

WESLEY, D. 1998. Life table analysis. *J Insur Med,* 30**,** 247-54.

WHITEHEAD, S. J. & ALI, S. 2010. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull,* 96**,** 5-21.

WHO 1952. *Constitution of the World Health Organisation. Handbook of basic documents,* Geneva: , Palais des Nations.

WHO 1999a. Difinition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Department of Noncommunicable Disease Surveillance,ed. Geneva: WHO.

WHO 1999b. WHO Consultation: Definition, diagnosis and classification of diabetesmellitus and its complications: report of a WHO consultation. Part 1:

Diagnosis and classification of diabetes mellitus. *WHO/NCD/NCS/99.2,; .* Geneva: World Health Organization.

WHO 2003. *Making choices in health: WHO guide to Cost-Effeciveness Analysis,* Geneva, World Health Organization.

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**,** 1047-53.

WILSON, N., ASHAWESH, K., SMITH, S. & ANWAR, A. 2008. The cost of screening for gestational diabetes mellitus. *J Med Screen,* 15**,** 213.

WU, S. Y., SAINFORT, F., TOMAR, R. H., TOLLIOS, J. L., FRYBACK, D. G., KLEIN, R. & KLEIN, B. E. 1998. Development and application of a model to estimate the impact of type 1 diabetes on health-related quality of life. *Diabetes Care,* 21**,** 725-31.

XIONG, X., DEMIANCZUK, N. N., SAUNDERS, L. D., WANG, F. L. & FRASER, W. D. 2002. Impact of preeclampsia and gestational hypertension on birth weight by gestational age. *Am J Epidemiol,* 155**,** 203-9.

XIONG, X., SAUNDERS, L. D., WANG, F. L. & DEMIANCZUK, N. N. 2001. Gestational diabetes mellitus: prevalence, risk factors, maternal and infant outcomes. *Int J Gynaecol Obstet,* 75**,** 221-8.

YUE, D. K., MOLYNEAUX, L. M., ROSS, G. P., CONSTANTINO, M. I., CHILD, A. G. & TURTLE, J. R. 1996. Why does ethnicity affect prevalence of gestational diabetes? The underwater volcano theory. *Diabet Med,* 13**,** 748-52.