A review and critique of modelling in prioritising and designing screening programmes

J Karnon, E Goyder, P Tappenden, S McPhie,
I Towers, J Brazier and J Madan

December 2007

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:
HTA Despatch   Email: orders@hta.ac.uk
c/o Direct Mail Works Ltd   Tel: 02392 492 000
4 Oakwood Business Centre   Fax: 02392 478 555
Downley, HAVANT PO9 2NP, UK   Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

A review and critique of modelling in prioritising and designing screening programmes

J Karnon,[*] E Goyder, P Tappenden, S McPhie,
I Towers, J Brazier and J Madan

School of Health and Related Research (ScHARR), University of
Sheffield, UK

* Corresponding author

Declared competing interests of authors: none

NIHR Health Technology Assessment Programme

The Health Technology Assessment (HTA) programme, now part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the costs, effectiveness and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The research findings from the HTA Programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA Programme is needs-led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, the public and consumer groups and professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA Programme then commissions the research by competitive tender.

Secondly, the HTA Programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Thirdly, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer-reviewed by a number of independent expert referees before publication in the widely read monograph series *Health Technology Assessment*.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

MR

Abstract

A review and critique of modelling in prioritising and designing screening programmes

J Karnon,[*] E Goyder, P Tappenden, S McPhie, I Towers, J Brazier and J Madan

School of Health and Related Research (ScHARR), University of Sheffield, UK

* Corresponding author

Objectives: To undertake a structured review and critical appraisal of methods for the model-based cost–utility analysis of screening programmes. Also to develop guidelines and an assessment checklist of good practice in the development of screening models.
Data sources: Major electronic databases of healthcare and operational research literatures were searched up to June 2003.
Review methods: Searches of the literature were undertaken to identify applied and methodological studies of economic evaluations of healthcare screening programmes. All applied screening models were also reviewed in three broad disease areas (cancer, cardiovascular disease and diabetes), as well as antenatal screening. A second-level review focused on particular aspects of the modelling process through case study assessments of screening models for three specific disease areas (colorectal cancer, abdominal aortic aneurysms and antenatal screening for haemoglobinopathies). A separate literature review of studies reporting the utility effects of screening was also undertaken. Guidelines and an assessment checklist for good practice for screening modelling were developed.
Results: Few relevant methodological studies were identified, and no studies reporting direct empirical comparisons of alternative methodologies were retrieved. From the review of disease-based screening models, it was apparent that many alternative modelling methods had been applied, including some relatively new approaches that had not been widely disseminated. Natural history modelling is the preferred approach. Alternative modelling approaches were generally only used to extrapolate the observed effects of screening and were unsuitable for evaluating unobserved screening options. More complex model structures may incorporate important additional aspects of the disease natural history, although any benefits should outweigh the consequences of additional unobservable input parameters and increased complexity in implementing the model. No direct comparisons of more detailed and less detailed screening model structures informed areas in which more realistic representations of the disease process may be most beneficial, so only general aspects of good practice could be defined. Two structural aspects that were not well handled by existing screening models included post-diagnosis disease progression and screening uptake. Most models described the former using historical mortality rates, rather than treatment models that are representative of current treatment patterns for different stages of the disease. Constant screening uptake rates were applied to all screening programmes and attendance was not linked to disease incidence or progression. Evidence exists to inform a more detailed representation of screening uptake. The most commonly applied modelling techniques were cohort Markov models and individual sampling simulation models. Individual sampling simulation models may provide more flexibility in their representation of a screening decision problem, but any benefits should outweigh the consequences of the need to assess both variability and uncertainty. Complex mathematical models describing input parameters as continuous variables have analysed the cost-effectiveness of screening; these require further development to estimate the cost–utility of screening directly, or to inform a more detailed representation of the preclinical section of a natural history model (with a traditional state-based model describing pathways' post-clinical presentation). Calibration is a common aspect of screening models, whereby models are fitted to observed data describing outputs of the model in order to populate unobserved input parameters. The review concluded that the estimation of a reference case input parameter set is not recommended.
Conclusions: The review of methods for the model-based cost–utility analysis of screening programmes identified the natural history modelling approach as the

preferred general method of evaluation for screening programmes. State transition models have generally been used to represent disease natural histories, with individual sampling models more prevalent than in treatment intervention evaluations. No comparative methodological studies were identified, so no empirical data were available to inform the relative merits of alternative methodologies. The defined guidelines and assessment checklist are informed, therefore, by theoretical interpretations of the impact of alternative approaches to different components of the modelling process when applied to the cost–utility analysis of screening programmes. Further research is needed into methods with the potential to improve the accuracy of screening models, and to respond to the needs of model users.

Contents

List of abbreviations

AAA	abdominal aortic aneurysm		IAOR	International Abstracts in Operational Research
ACE	angiotensin-converting enzyme		ICER	incremental cost-effectiveness ratio
ADPKD	autosomal dominant polycystic kidney disease		IGT	impaired glucose tolerance
AIC	Akaike information criterion		INFORMS	Institute for Operations Research and Management Sciences
CF	cystic fibrosis		LMP	last menopause
CHD	coronary heart disease		LSIL	low-grade squamous intraepithelial lesion
CI	confidence interval			
CIN	cervical intraepithelial neoplasia		MCH	mean corpuscular haemoglobin
CRC	colorectal cancer		MCMC	Markov chain Monte Carlo
CVS	chorionic villus sampling		MI	myocardial infarction
DALY	disability-adjusted life-year		MISCAN	Microsimulation for Screening Analysis
DCIS	ductal carcinoma *in situ*		NICE	National Institute for Health and Clinical Excellence
DES	discrete event simulation			
DVT	deep venous thrombosis		NPV	negative predictive value
EVI	expected value of information		NSC	National Screening Committee
FOBT	faecal occult blood testing		NT	nuchal transparency
FSIG	flexible sigmoidoscopy		ONS	Office for National Statistics
GDM	gestational diabetes mellitus		OR	operational research
Hb	haemoglobin		OTA	Office of Technology Assessment
HIV	human immunodeficiency virus		PCDP	preclinical detectable phase
HPV	human papilloma virus		PET	positron emission tomography
HRQoL	health-related quality of life		PND	prenatal diagnosis
HSIL	high-grade squamous intraepithelial lesion		POST	patient-orientated simulation technique
HYE	healthy-year equivalent		QALY	quality-adjusted life-year

continued

List of abbreviations *continued*

RCT	randomised controlled trial		SIL	squamous intraepithelial lesion
ROC	receiver operating characteristic		STAI	State–Trait Anxiety Inventory
SCD	sickle cell disorder		TOP	termination of pregnancy
SF-36	Short Form with 36 Items		TTO	time trade-off
SG	standard gamble		VAS	visual analogue scale

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Objectives

The objective of this report was to undertake a structured review and critical appraisal of methods for the model-based cost–utility analysis of screening programmes. It also aimed to develop guidelines and an assessment checklist of good practice in the development of screening models.

Background

Screening aims to identify disease or risk factors for disease prior to the point of clinical presentation in order to intervene more effectively. There are usually multiple options for the provision of screening, comprising alternative screening tests, eligible populations and screening intervals. Clinical trial data comparing a limited number of screening options over a relatively short time horizon may be available, but it is unlikely that trial evidence will inform all policy relevant specifications of a screening programme. Decision analytic models provide a structure for the explicit synthesis of information from a wide range of sources, as well as describing the uncertainty around the baseline results.

There are guidelines around the conduct of modelling studies in healthcare generally, but there are unique issues around the modelled evaluation of screening programmes that could be usefully informed by a thorough methodological review in this area.

Methods

Searches of the healthcare and operational research literatures were undertaken to identify applied and methodological studies of economic evaluations of healthcare screening programmes. In addition to identified methodological studies, all applied screening models were reviewed in three broad disease areas – cancer, cardiovascular disease and diabetes – and antenatal screening. This first-level review assessed broad issues such as the choice of modelling technique and general approaches to populating screening models.

A second-level review focused on particular aspects of the modelling process through case study assessments of screening models for three specific disease areas – colorectal cancer, abdominal aortic aneurysms and antenatal screening for haemoglobinopathies. A separate literature review of studies reporting the utility effects of screening was also undertaken.

The final stage of the review involved the development of guidelines and an assessment checklist for good practice for the conduct of model-based cost–utility analyses of screening programmes.

Results

Few relevant methodological studies were identified, and no studies reporting direct empirical comparisons of alternative methodologies were retrieved. Models for the evaluation of screening from outside the health field were found to have limited applicability to the evaluation of health-based screening, although a key set of papers were identified in the operational research databases. From the review of disease-based screening models, it was apparent that many alternative modelling methods had been applied, including some relatively new approaches that had not been widely disseminated.

Natural history modelling is the preferred general approach. These models describe disease progression from the point at which disease becomes detectable to death. When a screening model is laid on top of the natural history model, the course of the natural history is altered through the detection of disease at an earlier stage. Alternative modelling approaches were generally only used to extrapolate the observed effects of screening and were unsuitable for evaluating unobserved screening options.

More complex model structures may incorporate important additional aspects of the disease natural history, although any benefits should outweigh the consequences of additional unobservable input parameters and increased complexity in

implementing the model. No direct comparisons of more detailed and less detailed screening model structures informed areas in which more realistic representations of the disease process may be most beneficial, so only general aspects of good practice could be defined.

Disease states at the point of clinical presentation should represent prognostic indicators that influence treatment choices and treatment effectiveness, for example, a breast cancer screening model may describe health states as a function of tumour size, nodal status, oestrogen receptor status and menopausal status. If discrete states are used, the categorisation should reflect the relationship between the prognostic indicator and treatment choices and treatment effectiveness.

Two structural aspects that were not well handled by existing screening models included post-diagnosis disease progression and screening uptake. Most models described the former using historical mortality rates, whereas the preferred approach would incorporate treatment models that are representative of current treatment patterns for different stages of the disease. Commonly, constant screening uptake rates were applied to all screening programmes and attendance was not linked to disease incidence or progression. Evidence exists to inform a more detailed representation of screening uptake.

The most commonly applied modelling techniques were cohort Markov models and individual sampling simulation models. Individual sampling simulation models may provide more flexibility in their representation of a screening decision problem, but any benefits should outweigh the consequences of the need to assess both variability and uncertainty.

More recently, complex mathematical models that describe input parameters as continuous variables have analysed the cost-effectiveness of screening. These models require further development to estimate the cost–utility of screening directly, or to inform a more detailed representation of the preclinical section of a natural history model (with a traditional state-based model describing pathways' post-clinical presentation).

The review assessed a range of approaches to the estimation of input parameters that are specific to screening models, including preclinical disease incidence and progression and screening test

characteristics, although few applied approaches were identified in other areas, such as estimating screening uptake and the utility effects of screening.

Calibration is a common aspect of screening models, whereby models are fitted to observed data describing outputs of the model in order to populate unobserved input parameters. The review concluded that the estimation of a reference case input parameter set is not recommended. A preferred calibration process involves predicting output parameters for a large number of input parameter sets, with the accuracy of each set's predictions represented as a weight. The main analysis of the model involves sampling a large number of input parameter sets according to the weights attached to each input parameter set, from which mean values and probability distributions of cost-effectiveness can be derived.

Conclusions

The review of methods for the model-based cost–utility analysis of screening programmes identified the natural history modelling approach as the preferred general method of evaluation for screening programmes. State transition models have generally been used to represent disease natural histories, with individual sampling models more prevalent than in treatment intervention evaluations. No comparative methodological studies were identified, so no empirical data were available to inform the relative merits of alternative methodologies. The defined guidelines and assessment checklist are informed, therefore, by theoretical interpretations of the impact of alternative approaches to different components of the modelling process when applied to the cost–utility analysis of screening programmes.

Recommendations for further research

More complex mathematical modelling approaches have great potential as an alternative or adjunct to state-based modelling techniques for the evaluation of the cost–utility of screening programmes. Research is needed into the development of such models for the full evaluation of the cost-effectiveness of screening, and also a hybrid formation in which such techniques may be best suited to modelling the preclinical phases of disease.

There is scope for developing more comprehensive and explicit methods for calibrating models, which describe correlations between input parameters.

Empirical estimates of differences in the mean and probabilistic outputs of less complex cohort Markov models and more complex individual sampling models, using the same data sources,

would be of interest. Such comparisons may inform general areas in which simplifying assumptions are justified.

The direct utility effects of screening are under-researched, and may have a significant effect on the estimated cost utility ratios. More primary screening studies should incorporate utility measurements in their protocol.

Chapter 1

Introduction

Background

The purpose of screening is to identify disease or risk factors for disease in its presymptomatic or preclinical stage. There are well established criteria for screening programmes,[1] which have been adapted by the UK National Screening Committee (NSC) to guide the provision of screening programmes in general, and also to inform the specification of accepted screening programmes. The criteria address four broad factors: the condition, the test, the treatment and the screening programme.

The condition, it is stated, must be important, and the natural history and epidemiology must be understood. The screening test should be simple, safe, precise and acceptable to the general population, and there should be a defined diagnostic process following a positive test. Treatment for screen-detected disease should lead to better outcomes than treatment provided at the point of clinical diagnosis. Regarding the screening programme as a whole, it is stated that plans for monitoring the programme should be defined, adequate staffing and facilities should be available to cope with expected demand and the programme should provide value for money, as compared with other areas of medical expenditure.

The last criterion states the need for screening to be cost-effective and, implicitly, if screening is cost-effective, that the most cost-effective form of screening should be implemented. Each of the preceding criteria describe factors that must be defined in order to estimate the cost-effectiveness of screening, or that will permit the confirmation of cost-effectiveness, that is, monitoring. The criteria recognise the need for cost-effectiveness to be defined in terms of a generic outcome measure to allow comparison with other areas of medical expenditure, which in practice requires the estimation of quality-adjusted life-years (QALYs).

The criteria also state that there must be evidence from high-quality randomised controlled trials (RCTs) that the screening programme is effective in reducing mortality or morbidity. Trial data

comparing a limited number of screening options over a relatively short time horizon may be available by the time that a policy decision is required for new screening programmes. However, it is unlikely that trial evidence will inform all policy relevant specifications of a screening programme, incorporating alternative eligible populations, combinations of screening tests, screening intervals and treatment options. Data required to inform the full evaluative process are likely to come from a range of sources.

Decision analytic models provide a structure for ordering and synthesising information from a wide range of sources, such as the incorporation of expert opinion alongside the results of primary and secondary data analyses. Models promote the explicit formulation of assumptions, such that all issues captured in the model are open to scrutiny by experts, from both clinical and economic perspectives, thus allowing key issues and uncertainties to be identified and addressed. The development and use of decision analytic models also provides an opportunity to explore the sensitivity of the results of an evaluation to variations in the assumptions that underpin the model. Sensitivity analysis helps highlight areas in which further research is likely to be most useful.

The role of modelling in the economic evaluation of healthcare interventions is now widely accepted, as recognised in the submission criteria for the National Institute for Health and Clinical Excellence (NICE). Existing studies describing guidelines or methods of good practice for the conduct of modelling studies in healthcare,[2–4] including a review of such guidelines,[5] provide generic advice that should be incorporated into models of screening programmes. However, there are distinct issues relating to the modelled evaluation of screening programmes that could be usefully informed by a thorough methodological review in this area, such as modelling preclinical disease progression, screening test characteristics, screening uptake and treatment effectiveness post-diagnosis. Although there have been numerous case studies, no thorough methodological reviews have been published in this area.

Objectives of this report

The objective of this report is to review and critique modelling methods used in the evaluation of the cost-effectiveness of screening, primarily defined as the incremental cost per QALY gained. The report also aims to produce guidelines for the development and assessment of such models to increase the appropriate use of modelling in the evaluation of screening.

Review methods

The following sections describe the sequential components of the review.

Literature search

The objective of the literature search was to identify applied modelling studies of the cost-effectiveness of healthcare screening and studies describing methodological issues around the modelled evaluation of screening programmes.

The initial search of the literature included detailed searches of MEDLINE, EMBASE, the economic evaluations databases produced by the NHS and the Office of Health Economics, and specialist operational research (OR) sources such as the Institute for Operations Research and the Management Sciences (INFORMS) database and International Abstracts in Operational Research (IAOR). Studies were also identified from discussion paper series and a general search of the Internet. The search terms included combinations of indexed and free-text terms that described the context (e.g. screening), the process (e.g. economic evaluation) and the method (e.g. modelling). Full details of the search strategy are presented in Appendix 1.

Additional studies were identified from the reference lists of the retrieved papers and from responses to the project website, which was advertised via the main health economics and OR email discussion lists. Visitors were asked for additions to the list of the screening studies identified from the initial literature searches.

Abstract review

The search strategy identified over 7000 references. Abstracts from all identified references were reviewed by two members of the research team (JK, EG). Full papers were retrieved if the abstract indicated that the study was an applied model-based cost-effectiveness analysis of a screening programme, or that the study addressed a methodological issue of relevance to the model-based evaluation of screening programmes. The review was not limited to cost–utility analyses of screening programmes because the estimation and application of utility values are only one aspect of the modelling process.

Abstracts labelled as applied were categorised according to the condition and population group evaluated. The screening categories were based on those defined by the NSC, which set individual screening programmes in the context of five population-based programmes: antenatal screening, child health screening, screening for men, screening for women and screening in old age. Within the screening for men and women groups, two broad disease categories are defined: cancer, and cardiovascular disease and diabetes. Applied abstracts that were included by only one reviewer were categorised and retrieved if the relevant category was included in the review process (see the next section).

A total of 513 references were identified as potential modelling studies from the abstract review of the healthcare databases. The abstracts are arranged in the following categories (numbers of abstracts are given in parentheses):

- antenatal screening (92)
- childhood screening (12)
- cancer screening (211)
- cardiac/diabetes (65)
- contagious diseases (74)
- gastric-related (13)
- miscellaneous (46).

Ninety-nine abstracts were retrieved from the searches of the two main OR databases. The 99 abstracts were reviewed and labelled as being applied and/or methodology studies: 43 were defined as solely applied studies, 41 were defined as solely methodological studies and 15 were defined as both applied and methodological. Of the 58 applied studies, 26 evaluated screening for contagious diseases, primarily human immunodeficiency virus (HIV) (21 studies), and of the remaining 32 applied studies, 14 of the references were conference abstracts. Fifteen of the remaining 18 applied studies were reviewed as they concerned screening for diseases that were included in the case studies, although four of the studies addressed issues relating to the service implementation of screening.[6–9]

Of the 56 studies identified from the OR databases that were labelled as methodological,

13 were conference abstracts. A range of potentially relevant studies were reviewed from the remaining methodological studies. Models for the evaluation of screening from outside the health field were found to have limited applicability to the evaluation of health-based screening, because the processes are fully observable, large quantities of data describing the pathways of the objects of interest are available and the pathways are generally less complex than those associated with disease progression. Non-uniform screening intervals may also be easily adopted in industrial screening processes, which are not feasible from a healthcare service implementation perspective.

A key set of papers by Baker were identified as part of the OR literature review.[10–14] These papers described an updated approach to modelling breast cancer screening programmes, which was investigated in detail as part of the current review. From the review of disease-based screening models, including those identified from the OR literature searches, it was apparent that many alternative modelling methods had been applied, including some relatively new approaches that had not been widely disseminated. The research team decided, therefore, to concentrate efforts on assessing and critiquing studies identified from the healthcare screening modelling literature.

Full paper review

A two-level review process was implemented. The first level involved the broad review of a wide range of identified modelling studies in three of the screening categories defined by the NSC – adult screening for cardiovascular disease and diabetes, adult screening for cancer and antenatal screening. Childhood screening was not included as a separate category as few modelling studies were identified in this area, while separate reviews of screening programmes for men, women and older persons were not undertaken because it was hypothesised that the characteristics of different disease areas would have a greater impact on modelling methods than differences in the characteristics of the eligible population for alternative screening programmes. Screening programmes for contagious diseases were excluded from the review because models for the evaluation of such programmes are subject to very different characteristics, which could not have been adequately addressed in the current review. A separate literature review of studies reporting the utility effects of screening was undertaken.

The first-level review was a 'light touch' review of a wide range of models in three broad categories (antenatal, cancer, and diabetes/cardiovascular disease), which informed the choice and application of more detailed reviews of a specific disease within each of the three broad disease areas. The first-level review identified the range of modelling approaches that had been used to evaluate the different forms of screening, identifying areas of consensus and highlighting alternative approaches to screening-specific modelling issues that could be investigated further. The categories included in the first-level review were chosen because they contained the largest number of modelling studies, and because the modelling characteristics of the omitted categories were felt to be adequately covered by the chosen categories.

The second level of the review involved a case study analysis of a single disease from each of the three screening categories included in the first-level review. The modelling studies in each of the chosen categories were reviewed in more detail to investigate specific modelling issues identified during the first-level review. The areas of screening for colorectal cancer, abdominal aortic aneurysms (AAAs) and antenatal screening for haemoglobinopathies were chosen. The choice of disease areas for the second-level review was primarily based on the current policy relevance of screening for the conditions, which was informed by discussions with the clinical and policy advisors to the project (see Acknowledgements).

Data extraction

The first-level review extracted information relating to the categories presented in *Table 1*, which informed the broad reviews of modelling studies undertaken in the three broad disease areas (cancer, cardiac/diabetes, and antenatal screening).

It was known *a priori* that most screening models are based on models of the natural history of the relevant disease. A key element of the review was to assess the relative benefits of alternative applied natural history models for the cost–utility analysis of screening programmes. This aspect of the review involved a comparison of alternative modelling techniques and how the optimal choice of model varies according to the characteristics of the screening programme being evaluated. Most natural history models of screening programmes for diseases that develop over time (i.e. not diseases that are the subject of antenatal or neonatal screening) include some unobservable input parameters, primarily describing the incidence and progression of preclinical disease

TABLE I *Data extraction form for preliminary review*

Disease area	Relevant issues
Disease incidence	How have they measured incidence? Have they measured symptomatic or asymptomatic incidence?
Disease detection	How has the stage of disease at detection (either screen or clinical-diagnosis) been determined: sources and methods?
Disease progression	How is disease progression described, both preclinical and postclinical: sources and methods?
Test characteristics	How are the test characteristics estimated: sources and methods?
Treatment effectiveness	How is treatment effectiveness at screen- or clinical-detection estimated?
Model type	What type of model is used, does it have any special features, e.g. modelling resource constraints or dynamic incidence/prevalence?
Outcome measurement	What outcome measure is used, any comments on screening-specific estimation methods?
Cost issues	Any comments on screening-specific estimation methods?
Uptake	Do differential uptake rates affect cost-effectiveness, by what mechanism, e.g. is risk or cost a function of uptake?
Comments	Any other comments?

and the probabilities of clinical presentation. Applied methods for estimating the value of unobservable parameters were assessed.

The critique of methods and assumptions used to model interrelationships between input parameters covered the modelling of the effectiveness of the screening test, which may be a function of the efficacy of the screening instrument, the experience of the personnel involved in the provision of screening (both clinical and non-clinical), and the service organisation of the screening programme. Organisational issues include management variables such as the effectiveness of the record keeping and procedures for identifying the relevant screening population (e.g. the methods used to identify high-risk groups), and also policy variables such as the relevant screening interval.

The evaluation of resource use associated with screening programmes involved two main issues: (1) the measurement of healthcare resources and (2) the extent to which financial effects on alternative stakeholders are incorporated and the methods used to measure these effects. Participation costs, such as travel costs and the productivity effects of time off work, may be significant in screening programmes due to the population-based nature of the intervention.

The focus of the review is the use of models to assess the cost–utility of screening programmes, so the critique of outcome measurement concentrates on methods used to estimate utilities relating to

the outcomes and processes of screening. Utility estimation covers the impact of the disease of interest, and also the effects of screening *per se*, including positive effects such as reassurance and the provision of information, in addition to negative effects such as increased anxiety as a result of being screened.

The second-level case study reviews addressed specific issues and involved fewer studies, such that no formal review process was defined.

Guidelines development

The final stage of the review involved the development of guidelines that identify preferred modelling approaches that may be dependent on the screening context. The guidelines were defined by discussions between the authors of this report, which were informed by the findings from the review, the application of the reported case study evaluations and discussions over the course of the review with health economists and operational researchers with experience of modelling screening programmes (see Acknowledgements).

Outline of subsequent chapters

The body of this report comprises groups of chapters relating to the three broad screening categories described earlier. Chapters 2, 3 and 4 review and critique models used to evaluate screening programmes for cancer. Chapter 2 presents a broad review of studies that evaluated the cost-effectiveness of screening programmes for

colorectal, breast and cervical cancers. This chapter compares the key assumptions made by different studies with respect to the structure of screening models, and the broad methods for populating these models, particularly for the estimation of unobservable parameters. Chapter 3 presents a preliminary analysis of the potential benefits of an updated analytic modelling approach that has recently been applied to the evaluation of breast cancer screening. Chapter 4 presents a case study evaluation of alternative screening programmes for colorectal cancer (CRC). This chapter concentrates primarily on methods for the populating screening models, both for observable and unobservable parameters.

Chapters 5 and 6 review screening models for cardiovascular disease and diabetes. Chapter 5 presents a general review of such screening models, focusing on the perspective of potential policy makers. Chapter 6 presents a detailed analysis of modelling studies of screening programmes for AAAs that assesses the relevance the policy questions addressed and the modelling assumptions and the adequacy of the information presented about the model.

Chapters 7 and 8 address issues around the modelled evaluation of antenatal screening programmes. Chapter 7 reviews identified models of antenatal screening for three diseases – haemoglobinopathies, Down's syndrome and HIV.

These cover screening programmes for autosomal recessive disorders that require knowledge of the carrier status of both parents, chromosomal abnormalities that incorporate characteristics of the mother and unborn child and infectious diseases that involve the identification of the condition of interest in the mother. Chapter 8 presents a case study evaluation of antenatal screening for haemoglobinopathies, the focus of which is a comparison of the relative benefits of decision tree models and analytical mathematical models to describe the short-time horizons involved in antenatal screening evaluations.

Chapter 9 reports the findings of a separate review of utility measurements of events that are specific to screening. The main area of interest is the direct effect of screening on individuals who are screened as a result of the four screening results (true- and false-positive and -negative results). The potential utility impact of screen detected early diagnosis of disease is also considered.

Chapter 10 combines the results of the previous chapters to produce a structured set of guidelines that address the components of the screening modelling process. Chapter 11 converts the guidelines into an assessment checklist for future model-based screening evaluations.

Chapter 12 presents the conclusions from this study and lists areas for further research.

Chapter 2
Choice of modelling technique: a case study of cancer screening models

Introduction

The literature search described in Chapter 1 found that model-based evaluations of screening programmes have most commonly been applied to cancer screening, primarily covering breast, cervical, colorectal, prostate and lung cancers. A range of mathematical and simulation modelling techniques have been used to evaluate cancer screening. The applied mathematical models include simple decision trees, cohort Markov models and more complex approaches that override the Markovian assumption of exponential transition rates between health states.

This chapter assesses the benefits of alternative modelling techniques in terms of the accuracy (reality) with which the disease process is described, the strength and justification of the assumptions made and the uncertainty around key input parameters. The following three sections review screening models in the areas of colorectal, breast and cervical cancer. The relative merits of identified modelling techniques and broad methods of populating these models are then discussed.

Background to the natural history of cancer

The cancer process begins with a mutation in an originator cell. This mutation allows the cell to begin reproducing itself, forming a tumour. As long as the tumour remains localised, the cancer is unlikely to threaten the life of its host. It becomes more life threatening if metastasis has occurred. This is where a cancer cell has detached itself from the tumour and established a secondary growth elsewhere in the body. Once this has occurred, it is much more difficult to eradicate the mutated cells; treatment has to be more aggressive, with increased side-effects and reduced chance of success. The longer the cancer has been present, the greater is the likelihood of metastases occurring. Perhaps more important, though, is the histological grade of the cancer. This is a measure of the degree of mutation that has occurred. Cancer cells that show a greater degree of abnormality tend to divide most frequently, and therefore expand quickly. Even when small, it is thought that tumours of a high histological grade are more likely to produce metastases.[15–17]

The purpose of screening is to identify cancers at an earlier stage in their development, so that treatment has a greater probability of eradicating the cancer cells. Models of cancer screening must differentiate sufficiently between cancer stages to capture the effects of early diagnosis, in addition to reflecting variability in the length of the latent period (the time between the cancer becoming detectable and the point at which a patient presents clinically) that means that a proportion of patients with cancer will die of other causes.

Model review

Stevenson[18] cites an analytic model published by Lincoln and Weiss[19] in 1964 as the first example of a mathematical model developed for purpose of analysing cancer screening, which was followed by a range of analytic cancer screening models over the next two decades. The first simulation model is referenced to Knox.[20] These earlier models are excluded from the review as they assessed particular aspects of a screening strategy; for example, Zelen and Feinleib[21] define the lead time as a measure of screening benefit. More recent applications of the same general approaches that estimate the cost–utility of screening are included.

Colorectal cancer
Identified screening models for CRC vary in their interpretation of key aspects of the natural history of CRC, including the extent to which CRC develops from non-invasive polyps, and the relevance of modelling both adenomatous polyps, which can become larger over time and may develop into cancer, and hyperplastic polyps, which do not increase in size and do not become cancerous. The reviewed models of CRC are described below. The order in which the models are reviewed is based on an approximate gradation of the complexity of the model structures, starting with the less complex models. The first two models

reviewed are less complex in terms of their model structure because they do not adopt a natural history modelling approach; all of the remaining models describe the natural history of CRC.

Sonnenberg and colleagues[22,23] use a simple model that describes five main states: non-compliance with screening, status after screening, status after colonoscopy, status after polypectomy and CRC. Patients remain in the 'status after …' states until the next screening-related intervention unless they develop cancer in the interim period. If the next screening intervention is refused, patients enter the non-compliance arm and have a risk of cancer equal to the age-specific incidence.

A rate of preventive efficacy is seemingly applied to all persons with polyps who undergo polypectomy, whereas the cases that progress to cancer following polypectomy are assumed to be detected earlier, leading to improved survival. No relationship between adenomas and cancer is specified; for example, it is not clear which cancer incidence rates are applied to persons with a negative screen result, a group that includes true-negative and false-negative results.

Gyrd-Hansen and colleagues[24,25] combine estimates of test sensitivity and average sojourn time with age-specific incidence rates in a simulation process to estimate the number of cancers detected at each screening round of 60 possible screening programmes comprising alternative combinations of eligible ages and screening intervals. It is assumed that 30% of screen-detected cancer patients survive due to early detection; this percentage is assumed to be constant across different screening intervals. The life-years gained from the early detection of cancer are estimated as the sum of the age-specific life expectancies for surviving individuals after adjusting for mean lead time.

The impact of non-invasive polyps is based on the cumulative risk of large polyps progressing to cancer, which is applied to the observed difference in the number of polyps detected to estimate the number of cancers avoided due to the detection of polyps. The relative risk of CRC for persons in whom a polyp is detected and followed up is applied to age-specific incidence rates, and these cancers are assumed to have the same mortality risk as screen-detected cancers.

Ladabaum and colleagues[26] evaluate the use of aspirin as an adjunct to CRC screening. The model structure describes only an adenomatous polyp state from which all polyps have a probability of progressing to cancer. A proportion of cancers are assumed to originate from polyps. Cancer is divided into local, regional and distant. Separate sensitivity rates for polyps and cancers were estimated for faecal occult blood testing (FOBT) and flexible sigmoidoscopy (FSIG), with the sensitivity of FSIG accounting for the proportion of lesions within reach. Patients with detected polyps undergo a more intensive surveillance programme, although it is not stated whether alternative recurrence rates are applied.

Shimbo[27] uses a Markov model describing the progression of persons without polyps or cancer to one of three states: hyperplastic polyp, adenomatous polyp or cancer. Cancer is divided into Dukes' stages A, B and C. The model does not explicitly describe polyp location, although the estimated sensitivity of FSIG is adjusted for the proportion of lesions within reach. Persons with detected polyps are assumed to rejoin the normal state and do not have alternative recurrence rates (this assumption is tested in the sensitivity analysis).

Vijan and colleagues[28] also use a Markov model, in which persons with a normal epithelium may progress to a polyp state, or they may progress directly to the local cancer state. Polyps either remain as polyps or they enter a premalignant dwelling state or the local cancer state without passing through the premalignant state. Patients either remain in the premalignant state or progress to local cancer. Separate local, regional and disseminated cancer states are described. Constant durations are modelled for local and regional cancer through the use of separate year 1 and year 2 local states to represent 2-year duration, and a single regional state to represent 1-year duration, that is, all persons remain in each state for one annual cycle on their way to the disseminated state. The model does not explicitly define alternative polyp locations, although the estimated sensitivity of FSIG was adjusted to account for the proportion of polyps and cancers that could be reached.

Alternative incidence rates and follow-up strategies are described for adenomatous and hyperplastic polyps. A single recurrence rate for adenomatous polyps is described, but not for hyperplastic polyps. Presumably, all hyperplastic polyps remain in the initial polyp state until death or detection. Vijan and colleagues also define the proportion of multiple polyps and those larger than 1 cm, which are used to define surveillance schedules following detection of polyps.

Frazier and colleagues[29] use a Markov model to describe progression between low- and high-risk polyps (the latter defined as being over 1 cm or containing villous histology), and three stages of cancer (local, regional and distant). The model structure implies that all cancers originate from polyps. The two sides of the colon (the distal or proximal colon) are modelled separately because FSIG can only visualise the distal colon. It appears that identical transition rates are applied to polyps originating in either colon. This model incorporates separate sensitivity rates for low-risk polyps, high-risk polyps and cancer, in addition to differential recurrence rates for low- and high-risk polyps following their detection and removal.

Wagner and colleagues[30] do not explicitly describe their model structure, although it is implied that the model defines the time required for a 5-mm adenoma to progress to colorectal cancer. Although it does not appear that separate classes of adenoma are modelled (i.e. all adenomas were assumed to originate as size 5 mm), polyps are defined as either those that will progress to cancer or those that will not (i.e. hyperplastic and adenomatous). A proportion of cancers originate as polyps, and a proportion of cancers are lifetime latent. Cancers are defined as either early- or late-stage (corresponding to Dukes' stages A and B, and C and D, respectively), with a constant duration in the early stage for non-lifetime latent cancers assumed (lifetime latent cancers remain in the early stages). This approach implies that the benefits of screening are derived solely from the prevention of late-stage cancers because treated early-stage cancers are assumed to be cured.

The US Office of Technology Assessment (OTA)[31] states that their model builds on the Wagner model.[30] The main difference is that the OTA model does not explicitly describe the proportion of lifetime latent cancers, and does present age-specific survival rates for early-stage cancer (in addition to late-stage). The model assumes that all persons with detected polyps die of other causes.

Neilson and Whynes[32] describe a semi-Markov process that requires patient-level simulations, which also builds on the model structure developed by Wagner and colleagues. The general model structure describes progression from healthy to adenoma to early asymptomatic cancer to late asymptomatic cancer. It is also stated that a proportion of cancers are assumed to occur directly from the healthy state. The text mentions the existence of 'true non-progressive polyps', which implies that separate adenomatous and hyperplastic polyp states were defined. The prevalence and incidence of lifetime latent cancers are specified as input parameters, indicating that this aspect of the model follows Wagner and colleagues more closely than the OTA model.

The primary difference between the Neilson and Whynes model and the Wagner model appears to be the use of a semi-Markovian sampling strategy to describe the transition probabilities between states. Instead of assuming constant transition probabilities between states, Neilson and Whynes sample the next state to which each person will move, and then sample a holding time in the current state.

Khandker and colleagues[33] use a dynamic state transition model. The basic model describes the incidence of hyperplastic and adenomatous polyps. Undetected adenomatous polyps progress to undetected cancer, which progresses to a treatment stage (either through screen-detection or clinical diagnosis). More detail is included within this basic structure, such that the complete model contains over 60 states incorporating polyp histology (size and stage of development), location (distal or proximal), age (5-year intervals) and cancer stage (local, regional, and distant, and number of years in each stage). Decision trees are appended to each of the basic model states to incorporate the extra detail, for example, at the end of each cycle a decision tree is solved for all patients in the adenomatous polyp state to describe the proportion of patients progressing to alternative states within the model as a function of the interactions between the screening tests and the characteristics of the adenomatous polyps (e.g. histology and location).

Tunnel states are used to apply time-in-state dependent transition probabilities from the polyp state (modelled as 10 states of 2-year duration), which allow the probability of progression to increase as a function of dwell time in the colon. Persons with detected polyps are transferred to a surveillance state from which increased recurrence rates are applied. Tunnel states are also used to describe the progression of patients with undiagnosed cancer, which allows the use of alternative probabilities of detection or disease progression as a function of time with cancer.

Eddy[34] uses a nine-state time-varying Markov model. States describe persons who have no diagnosis of cancer, who have cancer diagnosed in various stages (A, B, C and D), who have died from cancer or who have died from other causes.

Additional complexity to this simple model structure is incorporated via the unique approach to populating a state-transition model, based on the calculation of the following five probabilities using a set of differential equations, solved by numerical integration:

- The probability that a person who has no diagnosis of cancer has an asymptomatic but potentially detectable cancer or non-invasive lesion incorporated age- and sex-specific cancer incidence rates, risk factors, the preclinical detectable phase (PCDP) and the history of previous screening tests.
- The likelihood that a screening test would detect an existing cancer or non-invasive lesions is a function of the lesion's stage of development, the random false-negative rate of the screening test(s), the location of the lesion and the region of the bowel reached by the screening test.
- The probability of alternative stages at detection is determined by the lesions rate of development, the history of previous screening tests and the random false-negative rate of the screening test(s).
- Stage-specific survival rates inform the probability of dying from cancer.
- Other cause mortality rates based on general population data.

Eddy[34] assumes that a set proportion of cancers emanate from polyps, and that a proportion of adenomas that reach 5 mm develop into invasive cancer.

Two patient-level simulation models were identified that incorporated significantly more detail than the models reviewed above. First, Loeve and colleagues[35,36] present a model of colorectal cancer screening, based on the general MIcrosimulation for SCreening ANalysis (MISCAN) approach reported by Habbema and colleagues.[37] The Monte Carlo simulation model consists of a disease and a screening part. The disease part simulates a large number of individual life histories, based on assumptions regarding the epidemiology and natural history of the disease. Then, a specified screening programme is applied to the life histories, which changes some of the histories, constituting the simulated effect of screening.

The colorectal MISCAN model describes the progression of separate lesions within individuals. Each individual can be assigned a risk index that describes their relative risk of developing polyps,

or the population can be split into strata that represent, for example, alternative risk categories based on population characteristics. Each polyp is assigned an anatomical site defined in terms of the part of the bowel and a percentage that indicates the localisation within this part. Progressive adenomas may progress through three size-related polyp states (<5, 6–9, >10 mm), and they may transit from either of the two larger polyp states to preclinical cancer. A proportion of non-progressive adenomas (purposefully not defined as hyperplastic polyps, which have separate consequences) remain in the middle polyp stage and the remaining non-progressive polyps end up in the largest polyp stage. It is assumed that all cancers originate from polyps, though this assumption is tested in the sensitivity analysis. Preclinical cancer can progress through four stages (I–IV) or become clinically diagnosed from any stage, from which point a stage-specific survival time is sampled. The modelling of the test characteristics incorporates a random element to false test results and a systemic factor, such that not all false results are independent.

A time of death from other causes is sampled for each individual, and this time is applied in the absence of a prior death due to CRC. Pathways between states can depend on age and the anatomical site. An intuitive aspect of the MISCAN model is that preclinical rates of progression through the cancer states may be correlated, which means that individuals who progress quickly from Stage I to II, are more likely to progress quickly through Stages II and III also, and vice versa. Similar correlations may be described for the preclinical progression rates through the polyp states (although polyp progression rates may not be correlated to cancer progression rates). This facility allows for the probability of clinical presentation to be defined as a function of a lesion's growth rate (as hypothesised by Baker for breast cancer[10]).

Ness and colleagues[38] use discrete event simulation (DES) to describe a similar model structure to Loeve[36] that follows the progression of separate adenomas within individuals. An individually assigned genetic risk value is used to adjust the age-dependent polyp incidence rate for each individual. Each polyp is described in terms of its size (<5, 6–9, >10 mm) and location (left, right, sigmoid, high-rectum, low-rectum). Incident polyps are assigned to either a 'fast-growing' or 'slow-growing' group. All polyps are assumed to progress to cancer, but many will not progress within the natural lifetime of the individual.

Polyps may progress to cancer from any of the three polyp states. Three cancer states are defined (local, regional and distant), and progression directly from local to distant cancer is permitted.

Both the MISCAN and the Ness models describe lifetime profiles for a large set of individuals and then apply alternative screening schedules to each life history that may alter the course of each person's lifetime events.

Breast cancer

Breast cancer probably has the simplest natural history pathway as there is only a single type of non-invasive cancer [ductal carcinoma *in situ* (DCIS)], which does not regress. Again, the reviewed studies are presented in order of approximate increasing complexity.

Both Allen and colleagues[39] and Mandelblatt[40] use decision trees to model individual screening rounds. End-points in the trees include healthy, local breast cancer, regional breast cancer or distant metastases. Separate mortality rates are described for the three breast cancer states.

Lindfors and Rosenquist[41,42] use a Markov model to describe the probability of any form of breast cancer being detected, surfacing as an interval cancer or presenting clinically in the absence of screening. An annual mortality rate is applied to all diagnosed cancers, with a relative risk reduction for cancer mortality being applied to screen detected and interval cancers in the screening group.

Salzmann and colleagues[43] use a Markov model that includes the following health states: healthy, alive with breast cancer, breast cancer death and death from other causes. It appears that the model assumes a similar incidence of breast cancer, regardless of the screening strategy, and applies a mortality reduction after a defined delay between the start of screening and the onset of the mortality reduction; for example, screening women aged over 50 years provides a 27% mortality reduction 5 years after the start of screening.

Lai[44] describes two separate Markov chain models that each describe two breast cancer states: regional lymph nodes negative (or tumour size smaller than 2 cm) and regional lymph nodes positive (or tumour size larger than 2 cm). All states may be either preclinical or clinical. Mortality rates are based on the state at diagnosis or detection.

The breast cancer MISCAN model describes progression though a non-invasive state (DCIS) and three discrete stages of invasive cancer based on the size of the tumour (<10, 10–19, >20 mm).[45,46] Tumours in each of these categories may be preclinical, screen-detected or clinically diagnosed. The duration of the preclinical phase is modelled as being age-specific, based on 10-year age groups. Breast cancer mortality rates are based on the size of tumour at the time of diagnosis or detection, although it is recognised that the inclusion of lymph node status would add to the completeness of the model.

The model presented by Szeto and Devlin[47] is based on the MISCAN breast cancer model, where the main difference is the handling of post-diagnosis/detection survival. Once diagnosed, patients are cured or not, with separate age-dependent survival times applied. Non-cured survival times are referenced from a log-normal model for survival.

The breast cancer screening model developed by Eddy[48] has a similarly simple structure to the corresponding colorectal cancer screening model.[34] Additional detail is incorporated through a set of equations describing the probability that women with cancer will be detected if screened, or will present with cancer between tests (as a function of risk factors and screening history). If a cancer is detected, a second set of equations calculate how early it was detected as a function of how it was detected and screening history. A third set of equations calculate the probability of dying from breast cancer at any point in time as a function of the earliness of detection.

Other mathematical models describe the natural history of breast cancer as a series of continuous variables. Schwartz[49,50] assumes that tumours grow at a continuous rate and that the involvement of lymph nodes is a function of the age of a tumour, although seven discrete tumour sizes are defined in order that differential sensitivity rates may be applied. The rate of clinical diagnosis (in the absence of screen detection) is modelled as a function of tumour age and the mortality rate is modelled as a function of discrete prognostic groups that account for tumour size and lymph node involvement.

Baker[10] lists a similar set of structural assumptions, although the rate of clinical diagnosis is stated to be a function of size and growth rate, as is the

breast cancer mortality rate. Potential structural relationships are also discussed, including the likelihood of a non-random element to sensitivity (i.e. the probabilities of detecting cancers at successive screens is not independent), and that the clinical diagnosis rate is also a function of the time since last screen.

Parmigiani[51,52] uses an individual sampling simulation model that is populated through the analytic estimation of probability densities for each of the transitions within the model. The model describes the simple transition of patients from healthy to preclinical disease to clinical breast cancer to death, with death occurring from any of the three previous states. However, the estimated transition probabilities incorporate a considerable amount of detail, which leads to the specification of a set of four prognostic factors at the point of diagnosis that are a function of age at transition to preclinical and time spent in the preclinical state. Survival rates post-diagnosis are determined by the combination of prognostic factors.

Cervical cancer

The detection of non-invasive lesions is a key objective of screening for cervical cancer. Most cervical cancer screening models describe progression through a series of non-invasive states, either defined on the basis of different grades of cervical intraepithelial neoplasia (CIN)[53,54] or cervical squamous intraepithelial lesions (SILs).[55,56] A distinction between the non-invasive phases of colorectal and cervical cancer is that colorectal polyps can be described as progressive or non-progressive, whereas cervical neoplasia are commonly modelled as progressive or regressive. Recent cervical screening models also model the development of non-invasive lesions as a function of the presence of human papilloma virus (HPV) in women.[55,57,58]

All of the studies cited in the previous paragraph used a cohort Markov model to describe the natural history of cervical cancer, incorporating similar model structures. All represent the possibility of disease regression from the non-invasive state, or the earlier non-invasive states if separate non-invasive states are modelled. The handling of survival post-diagnosis varies between these studies; for example, Fahs[54] describes two invasive states, and applies age- and stage-specific mortality rates, whereas Maxwell and colleagues describe four invasive states,[58] but apply only stage-specific mortality rates. The endpoint in Jenkins and colleagues' model is invasive cancer.[53]

Knox also used a cohort Markov model,[20] which enabled transition probabilities to be specified as a function of age or age at onset of cancer.

Eddy[59] applied his generic Markov model approach to evaluate cervical cancer screening. It is not clear how the progression from non-invasive lesions to cancer is handled, although regression from non-invasive states is recognised. The model describes a total duration of the preclinical phase that is not age-specific, but that does differentiate between two types of lesion, which may be either fast- or slow-growing. Four stages of invasive cancer are described, and stage-specific mortality rates are applied. A constant 'random' false-negative rate is assumed, referring to the fact that some cancers will be consistently missed due to biological features of a lesion.

Non-Markov models include the cervical cancer MISCAN model,[60,61] which describes a single, non-invasive state and three cancer states, which allows for regression from the non-invasive state. Mortality rates are based on the age and stage at diagnosis. Separate test sensitivity rates are specified for the non-invasive and cancer stages.

Parkin[62] also used a microsimulation approach in which individual life histories are generated, to which alternative screening policies are applied. However, the life histories are generated by a Markov model based on the Knox[20] model. Transition probabilities within the natural history model are age-specific, using varying categories of age for the different transitions. Transition probabilities from invasive cancer to death are also dependent on current duration in the invasive cancer state. At the end of each cycle (year), each patient passes through the microsimulation model that describes the impact of the screening programme on each individual. The screening model is defined by the individuals to whom screening is offered each year (as a function of age, marital status, probability of childbirth and time since last screen), attendance rates and test characteristics.

Gustafsson and Adami[63] used an identification technique to model the natural history of cervical cancer. Their model describes progression from healthy to *in situ* to preclinical invasive to clinical invasive to death; regression from *in situ* to healthy is possible. Simulation is used to solve a set of differential equations. Progression and regression rates from carcinoma *in situ*, and dwelling times for *in situ* and preclinical invasive cancer are

estimated for separate birth cohorts by age (so one can see if age-specific rates vary by birth cohort). It appears that mortality rates from the single clinical invasive state are also estimated for separate birth cohorts by age.

A follow-up paper modelled the optimisation of screening intervals for cervical cancer using the same model structure in which a set of arbitrary rates of screening efficiency are tested.[64] Screening efficiency is defined as a function of screening attendance, test sensitivity, and completeness of diagnostic work-up.

Barton and Byran[65] used discrete event simulation to model capacity of reading laboratories, such that new screens enter a queue. The model structure is based on the Myers model,[55] which follows patients through HPV, two non-invasive states [low-grade SIL (LSIL) and high-grade SIL (HSIL)], and four invasive states (Stages 1–4). Regression is modelled from LSIL and HPV. The capacity of the reading laboratory is set and the minimum time per smear is estimated as 1 year divided by the annual capacity, that is, each presenting smear takes this long to be read. If demand outstrips supply a queue builds up and a woman's health state could progress during the queuing period.

Comparison of model structures and assumptions

This section assesses issues around the choice of model structure for the evaluation of screening programmes. One screening model approach is to model the incidence of cancer and apply trial-based risk reductions to cancer mortality rates to describe the benefits of screening.[43] The main disadvantage of the 'mortality reduction' approach is that it requires the application of observed mortality reductions, which precludes the evaluation of screening programme configurations (e.g. combinations of screening tests and intervals) for which there are no observed estimates of effect. The mortality reduction studies lack clarity, which may be due to the estimation of costs based on the screening process in the screened population, and the estimation of effects on the basis of cancer incidence in the non-screened population.

Decision tree models are overly simplistic and are also restricted to observed screening programmes.[39,40] These models describe alternative stage distributions of cancer at diagnosis, to which separate mortality rates are applied.

The majority of the reviewed screening models describe the natural history of cancer in the absence of screening. The impact of screening is then described by overlaying alternative screening programmes on top of the natural history. This general modelling approach incorporates data from screening trials but is not restricted to screening options that have been directly evaluated.

Van Oortmarssen and colleagues[45,46] describe a general model of the natural history of cancer and its interaction with screening, as reproduced in *Figure 1*. The model defines two broad cancer states: non-invasive and invasive. Non-invasive disease may regress, such that individuals may re-enter the 'no cancer' state. The state in which cancer is diagnosed (invasive or non-invasive), in combination with the mode of detection for invasive cancer (screen or clinical), affects the prognosis of the patient. The general model also assumes that a cancer may be defined as cured. This is a modelling artefact; as it is not possible to define a cured cancer patient following treatment, this state includes all persons who are diagnosed with cancer who do not die from cancer.

This basic model structure describes the underlying framework for all of the identified natural history models, although three broad modelling techniques have been used to represent the structure: cohort Markov models, individual sampling simulation models and more complex mathematical models that override the Markovian assumption.

The cohort Markov model is the standard technique used to model the economic impact of healthcare interventions over time, and is also the most common approach identified in the review. Events are modelled as transitions from one health state to another. The time horizon covered by the model is split into cycles of equal length. At the end of each cycle, the cohort of individuals may move to a consequent health state, or remain in the same state (unless the current state is a tunnel state). This process of moving between states continues until a patient enters an absorbing state, such as the state 'dead'. Transition probabilities are conditional on the current health state, but they may also vary according to the overall time spent in the model.

Individual sampling simulation models allow greater flexibility. An example of such a technique is DES, which is an event-orientated modelling approach. A DES model asks what and when is the next event for every person at the point at which

13
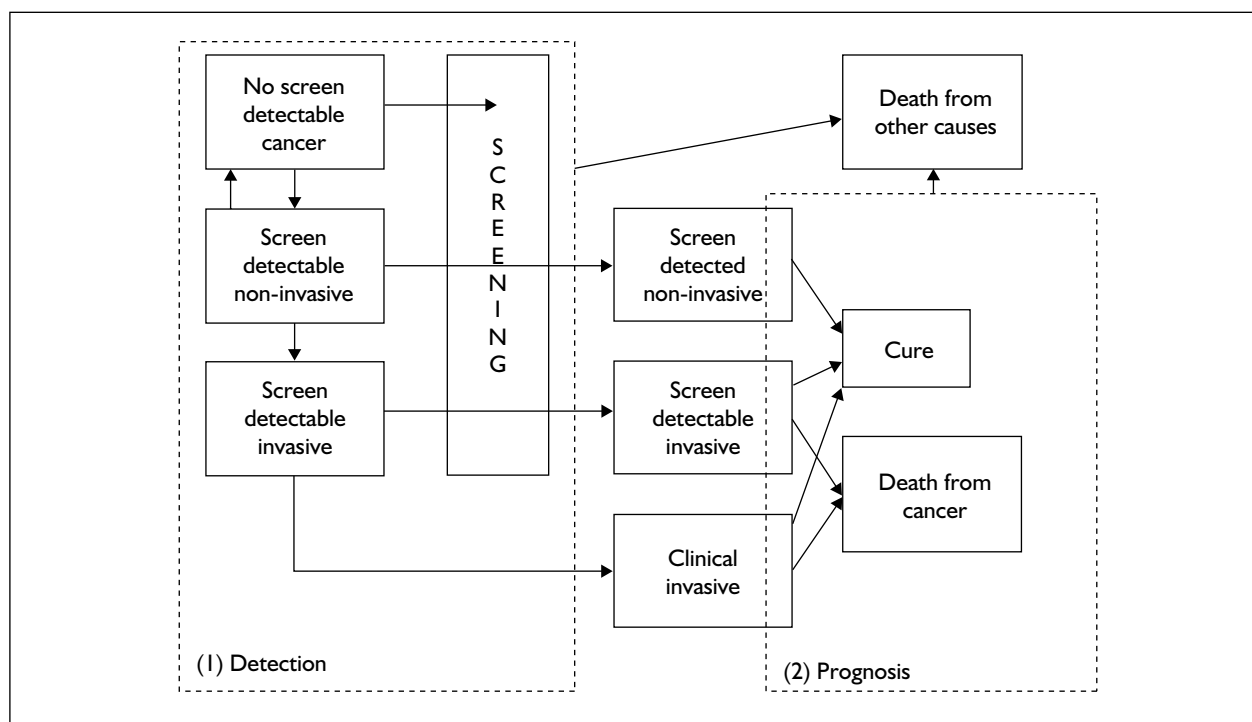
FIGURE 1 *General cancer screening model*[45,46]

they experience their current event, rather than a Markov model, which asks what events are occurring at regular intervals. DES models use attributes that record relevant elements of each individual's pathway through the model and personal characteristics, which can then influence future pathways, such as duration of the preclinical period for each patient. By following individuals through the model, it is also possible to assign individual state durations sampled from any form of probability distribution. DES models may only be analysed using individual-level (first-order) Monte Carlo simulation, which increases the analysis time.

The representation of the natural history is noticeably less complex in the reviewed Markov models, such as in the application of age-specific transition probabilities. Markov models tend not to facilitate the description of age-specific values for parameters such as the duration of the preclinical states and mortality rates from the point of cancer diagnosis. Only Knox[20] described age-specific durations of the preclinical phase in a Markov model (of cervical cancer screening), which includes 26 separate health states. Non-age-specific transition probabilities may be a reasonable assumption based on data and informed opinions about the natural history of the diseases, although the inclusion of age-specific

preclinical disease progression rates and mortality rates in other, less restrictive, models indicates that there is some evidence to support their incorporation. Differential age-specific preclinical disease progression rates are likely to affect the cost-effectiveness of screening at different ages, for example, if younger age groups have a shorter mean preclinical period then, *ceteris paribus*, shorter screening intervals will be more cost-effective at younger ages.

The assumption of constant transition probabilities means, for example, that an individual who has just developed a polyp has the same probability of progressing to cancer as an individual who has remained in the polyp state for 20 years. In a typical evaluation of a treatment intervention, for the same population mean duration in a state, the impact of modelling the wrong transition profile across the population would be limited to a slight error in the discounted outputs from the model (costs and effects).[66] However, in a screening model, the potential error is increased because the screening programme is laid on top of the natural history model, and an inaccurate transition profile will affect the estimated effectiveness of the screening programme. As an example, *Figure 2* presents a Weibull distribution of the probability of disease progression as a function of time (solid curve), and
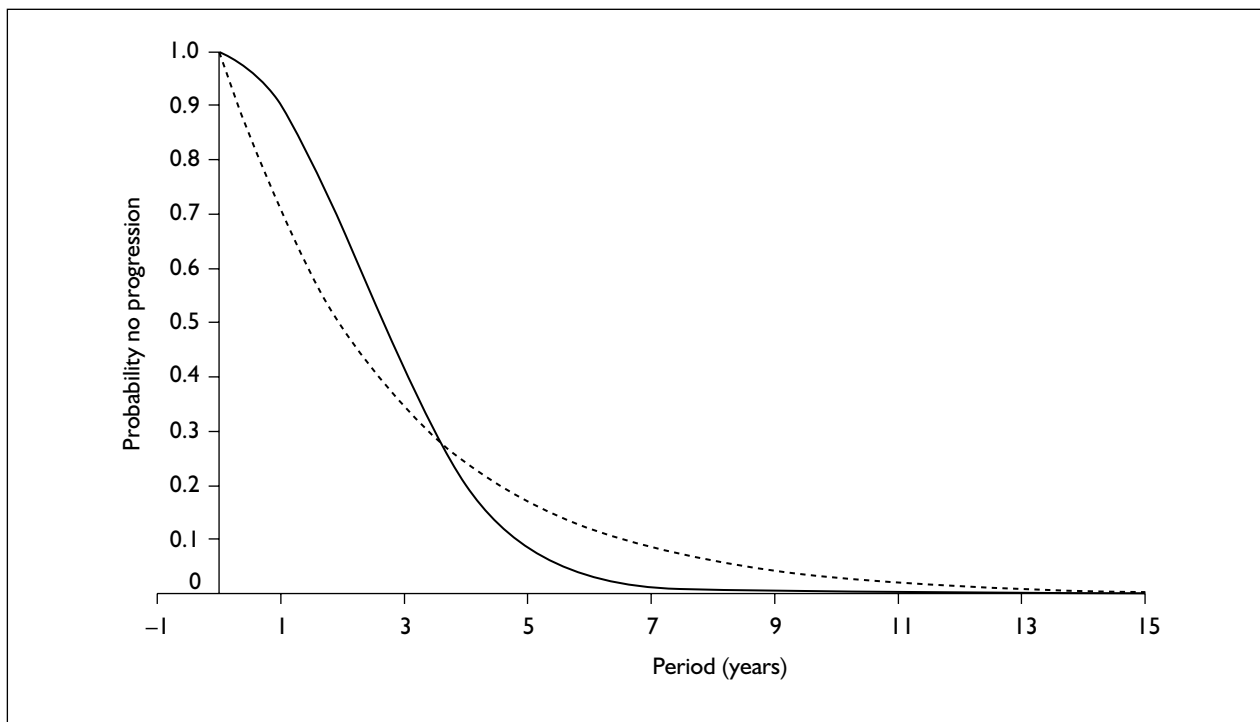
FIGURE 2 *Demonstration of constant and time-dependent progression rates*

an exponential curve with the same mean time to disease progression (dashed curve). The standard Markov approach assumes a constant probability of progression as represented by the exponential curve. If the time variant distribution is appropriate, the application of a constant progression probability clearly overestimates the early rate of progression, which means that the model would favour shorter screening intervals in the early period. Using tunnel states that describe, for example, alternative annual probabilities of progression, the cohort Markov approach can apply time varying progression probabilities that better reflect the distribution of the progression times. Shorter tunnel states will better approximate the patient level approach, though if screening intervals are defined by integer years (e.g. 1-, 2- or 3-year intervals) there are no real benefits to defining tunnel states covering less than 1-year intervals.

The representation of time-in-state dependent probabilities is only one modelling characteristic that requires additional states using a cohort Markov approach. Other 'health state multipliers' include the representation of:

- previous clinical events on future transitions, such as higher incidence rates following the removal of precancerous lesions

- individual attributes on future transitions, such as risk factors for the disease that may be correlated with screening uptake
- separate disease components within individuals, such as the development and progression of multiple cancerous lesions.

CRC may involve the most potentially complicated model structure due to the clinical relevance of multiple lesions. It is not feasible for a cohort-based Markov model to describe the progress of all lesions that may develop in both sides of the bowel, but it may be feasible to model the pathways of the most developed lesions in each side of the bowel. Combination health states can be modelled, such as 'low-risk proximal polyp and high-risk distal polyp' and apply transition probabilities to subsequent states, 'low-risk proximal polyp and distal cancer stage A' and 'high-risk proximal polyp and high-risk distal polyp', as described in *Figure 3*.

The representation of this level of detail soon leads to an overly complex model. A CRC screening model that incorporated two polyp states (low- and high-risk) and four cancer states (A, B, C and D), in theory would require 48 separate health states to model all single and multiple lesion states. If time-dependent transition probabilities are also represented using tunnel
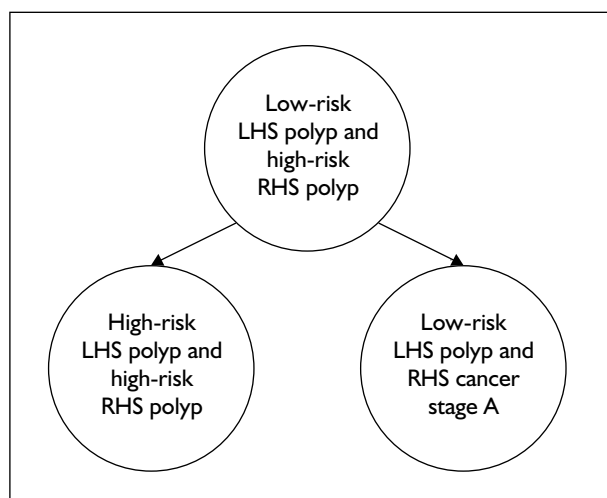
FIGURE 3 *Representation of a complex natural history using a cohort-based state transition model*

states, then the number of states included in the model will be measured in the hundreds.

The simplifying assumption that a proportion of proximal lesions are associated with 'sentinel' distal polyps that leads to the subsequent colonoscopic detection of proximal lesions is made in many of the reviewed cohort Markov models for CRC screening. Individual sampling simulation models describe the progression of separate lesions as attributes of the persons in the model,[35,36,38] which is a more intuitive approach and which estimates the proportion of sentinel polyps as a model output, rather than as an input parameter. Additional output parameters provide a wider basis for the calibration or validation of a model, though the simplifying assumption may be considered reasonable in order to maintain the benefits of the cohort Markov approach.

Loeve and colleagues[67] investigated the impact of systematic false-negative results, under the hypothesis that a proportion of lesions will remain undetected by FOBT because they will never bleed. To model the existence of systematic false-negative lesions in a cohort Markov model approach would require the specification of separate sets of health states to describe the progression of lesions to which a non-systematic, and a systematic, false-negative rate could be applied, that is, a doubling of the preclinical health states. Using the individual sampling approach, Loeve and colleagues were able to specify the nature of a lesion as an attribute that determined the appropriate false-negative rate for each lesion.

An additional issue common to both the Markov models and individual sampling simulation models identified in the review concerns the level of detail used to model post-diagnosis survival. Only one model identified across all cancer screening models incorporated a treatment model from the point of diagnosis.[68] Most cancer screening models applied a stage-specific survival period from the point of diagnosis, to which mean cost and utility values were applied. Post-diagnosis treatment models facilitate predictive models of survival that are based on current (or future) treatment protocols, whereas observational studies will likely be based on old management pathways that may not reflect current practice.

The most accurate screening model structure, incorporating the strongest possible set of assumptions, may be facilitated by more complex mathematical models. These models comprise a series of related differential equations that incorporate the assumptions made about the natural history (e.g. that there is a distribution of tumour growth rates), screening (e.g. sensitivity is a function of tumour size) and possibly interactions between disease progression and screening (e.g. the timing of presentation with clinical symptoms is a function of the time since last screen). The equations are solved to define probability distributions for each output parameter, for example, tumour sizes for screen-detected and clinically diagnosed cancers, or survival times for screened and non-screened cohorts.

More complex mathematical models describe input parameters as continuous variables that change smoothly over time, which means that parameters that are functions of other parameters, such as age-specific event rates, are modelled to their most exact specification. In a state transition model, applying the same stage-specific mortality rate to screen-detected and clinical diagnosed patients will bias results against screening because screen-detected cases will, on average, be detected earlier within a stage. The estimation of mortality rates as a continuous function of the stage at diagnosis is more likely to overcome this potential source of bias.

Most of the complex mathematical models that have been applied to cancer screening have not estimated the cost utility of screening. Baker,[10] for example, minimises a cost function to identify the optimal screening programme, in which cost values are applied to each screening test undertaken and to each month of life lost.

Parmigiani[51,52] uses a hybrid approach in which probability densities describing transition rates between health states are derived using a range of complex mathematical techniques, which allows the standard application of cost and utility values to defined health states along person pathways. Output distributions, such as the set of distributions describing the duration of the preclinical phase for women developing breast cancer at different ages, are combined to calculate the number of women transiting from the preclinical to the clinical stage at any point in time. This analytic approach differs from the Baker approach as the latter estimates the parameters for the set of equations that describe the full screening model simultaneously, whereas Parmigiani estimates different sections of the model separately.

General parameterisation approaches

This section assesses the merits of alternative general approaches for the population of screening models. Three broad parameterisation approaches can be defined from the review of cancer screening models: the non-calibration, the partial calibration and the complete calibration approaches. The non-calibration approach describes the independent estimation of all input parameters. This is the standard approach to decision analytic modelling for most treatment-based evaluations, although the existence of unobservable screening parameters such as progression between preclinical cancer states and the probabilities of clinical presentation make the non-calibration approach less common in screening models.

Examples of direct estimation methods of unobservable parameters include Parmigiani, who solved an integral equation for the unobservable incidence rate of preclinical tumours that included previously developed theories of the growth rate of breast tumours.[51,52] Alternatively, the non-calibration approach may rely on expert elicitation to populate the unobservable parameters. The MISCAN colorectal model was populated over the course of a 2-day expert meeting at which all input parameters were defined.[35,36]

Given the uncertainties around the non-calibrated estimation of unobservable parameters, the process of validation is crucial. Validation is preferably undertaken against output parameters derived from the defined eligible population for the screening programme being evaluated. Such data may include age-specific screen and clinical diagnosis rates and cancer mortality rates. Initial model specifications may result in poor representations of observed data, which requires 'reconsideration' of original parameter estimates. This process may be interpreted as calibration, although a key advantage of starting with a full set of input parameters is that it provides a grounded basis for each parameter estimate and an iterative process for re-estimation, which may reduce the problems of identifying the most likely baseline parameter set.

The partial calibration approach is perhaps the most commonly implemented approach to the population of state transition cancer screening models, which involves estimating the observable input parameters directly, and then calibrating the remaining unobservable parameters. The initial MISCAN breast cancer model treated the duration of the preclinical phase (split into two states) and the sensitivity of the screening test in these two states as unobservable.[46] The best fitting point estimates for these parameters were presented, and also an 'area of combinations' of the duration of the preclinical phase and average sensitivity that were in agreement with the results of the observed study.

In the original description of the MISCAN approach, Habbema and colleagues stated that, ideally, the goodness-of-fit would be tested against a series of implemented screening programmes, which would allow the best combination of assumed data inputs to be chosen.[37] Van Oortmarssen and colleagues implemented this approach using the breast cancer MISCAN model, comparing model predictions with observed data from two alternative Dutch screening programmes.[45] Parameter estimates calibrated to a previous trial (the HIP trial) resulted in a poor fit, so input parameters were varied systematically to obtain "an adequate overall fit of almost all screening results" from the two screening programmes. Reasons for the lack of fit included that aspects of the implementation of one of the screening programmes were not captured by the structure of the MISCAN model, such as the non-dichotomous interpretation of screening results and the use of alternative screening procedures at repeat screening rounds. It is implicitly assumed that the epidemiology and natural history of breast cancer are similar in the two populations, which may be reasonable as both trials were undertaken in The Netherlands within a similar time frame. However, if trials covering divergent populations and time

frames are used, it may be necessary for the cross-validation process to consider factors that are most likely to stay constant across different populations. Genetic, environmental and birth cohort effects will affect incidence rates, but how do such factors affect growth rates?

The full calibration approach is demonstrated by Baker,[10] who used maximum likelihood estimation to calibrate her breast cancer model to datasets describing tumour size for screen-detected cancers, the timing of presentation and tumour size for interval cancers, survival times for screen- and clinically detected cancers, prescreening age-specific cancer incidence and all-cause mortality rates. The likelihood function "is a complicated function of data and model parameters that for this problem requires numerical evaluation of double and triple integrals" (pp. 103–4). An objective measure – the Akaike information criterion (AIC) – of the best-fitting parameterisation is minimised to estimate the baseline set of input parameter values. Baker has published a separate paper that describes a form of sensitivity analysis for models fitted to data by statistical methods, which uses the covariance matrix of the fitted model parameters.[11]

Intuitively, the full calibration method may be criticised for not incorporating the full range of available data informing the observable input parameters, for example, the wide range of independent studies that have investigated the test characteristics of alternative screening tests. However, the output data used to calibrate the model may be more representative of the eligible population; for example, Baker used data sources from north-west England, and the comparability of the datasets was good because the non-screened cancer incidence rates from 1987 were combined with screened incidence rates from 1988 to 1990.[10] Ideally, the full calibration method should be validated using output data from a range of alternative populations, preferably within similar geographical locations and time frames. As described for the partial calibration approach, rational explanations for variations between populations should be sought and provided.

Discussion

This chapter has reviewed modelling studies evaluating screening programmes for breast, cervical and colorectal cancer. A summary of the applied modelling approaches provided a background to a comparison of alternative modelling techniques with respect to model structure and broad approaches to populating screening models.

That more complex modelling techniques facilitate a more accurate representation of the process being modelled is no surprise, but models are intended to simplify reality to represent the elements of a process that affect the relevant model outputs, which means that the less complex models are not necessarily inappropriate. However, the additional flexibility of more complex modelling techniques may be more likely to produce alternative policy conclusions in an evaluation of screening programmes than in a treatment-based health technology assessment due to the time-dependent relationship between the natural history of a disease and the implementation of screening.

In theory, adaptations to the implementation of the cohort Markov approach could address many of the observed areas of additional complexity, but the feasibility of building Markov models to the size required to incorporate such extra detail must be questioned. Individual sampling models facilitate the more intuitive representation of the natural history of cancer, but they can only be analysed using first-order Monte Carlo simulation, which increases the time required to build, parameterise, verify, validate and analyse the model, even more so if probabilistic sensitivity analyses and expected value of information (EVI) analyses are to be undertaken.

The third general modelling technique is termed the 'complex mathematical modelling' approach. This is the least developed of the techniques with respect to the estimation of the cost-effectiveness of screening. Both Eddy[34,48,59] and Parmigiani[51,52] used analytic approaches to estimate parameter values for state transition models, but only Baker[10] attempted to estimate the cost-effectiveness of screening directly using an analytic approach. The Baker model incorporates a high level of detail with respect to the disease process and its interaction with screening, primarily through the representation of model parameters as continuous variables. However, the estimation of the relative costs of alternative screening programmes was limited and further research is required to develop this potentially valuable methodology. Some further analysis of this general approach is presented in Chapter 3.

The obvious difficulty is in establishing the (in)significance of simplifying assumptions. In

practice, such exclusions are justified (if at all) on the basis of a lack of data or the opinions of experts. More appropriate would be a discussion of the likely direction of broadening the assumptions, that is, which screening programmes would become more or less cost-effective. If possible, an assessment of the magnitude of the impact should also be included. If a reasonable case can be made that a more complex representation of reality would not alter the conclusions drawn from the analysis, then a less complex model may suffice. Available evidence around the areas in which simplifying assumptions have been made should be collated to inform discussion around the appropriateness of the model. Chapters 5 and 6, which cover the presentation of screening models, provide practical examples of how the above issues can be addressed.

The main issue around the common, partial calibration, approach to populating screening models concerns the identification of input parameter point estimates when a wide range of calibrated parameter values provide a similar goodness-of-fit – should the baseline analysis use the combination that maximises the goodness of fit, or the midpoint of the area of combinations that are within a defined threshold of goodness-of-fit? Church suggests that as more than one combination of parameter estimates could result in a similar goodness-of-fit, a full analysis of the uncertainty around the combinations of the fitted parameters is necessary.[69] Probabilistic sensitivity analyses could be undertaken using probability distributions based on the 'area of combinations'. This approach was not identified in the review, although it is applied in the case study evaluation reported in Chapter 4.

Conclusions

This chapter has highlighted specific areas within natural history-based cancer screening models in which the level of detail facilitated by three broad alternative modelling techniques differs. No empirical evidence of the impact of these differences on resulting estimates of cost-effectiveness was identified, although there appears to be greater potential for differences due to the interaction between the screening programme and the natural history.

Alternative approaches to populating screening models should be judged on the merits of their application; for example, are the data sources relevant and appropriately applied, and has the choice of parameter values been cross-validated?

19

Chapter 3

Preliminary assessment of complex mathematical modelling methods for the cost–utility analysis of screening

Introduction

The previous chapter reviewed models that have been used to evaluate cancer screening programmes. The majority of the identified studies employed state transition models that used simulation methods to calibrate unknown input parameters to observable model outputs. A less common alternative approach used complex mathematical functions to estimate continuous disease functions. This chapter describes two applications of the alternative approach by Baker[10] and Parmigiani[51,52] in some detail and discusses the potential for their extended use in the economic evaluation of screening programmes.

The Baker model

Baker[10] uses an analytic, variable-based model to evaluate breast cancer screening. Baker starts by setting out a list of assumptions, derived from existing knowledge of breast cancer, which are used to develop the model structure (a set of variables, and mathematical equations representing the relationship between them). The variables are tumour size, age at origination, growth rate, presentation (in the absence of screening), metastasis, survival (with and without cancer) and mammographic sensitivity. These variables are characterised by probability distributions, which may be conditioned on time, or on one of the other model parameters. In Baker's model, presentation and screen sensitivity are conditioned on tumour size. This is intuitively plausible, as a larger tumour is more likely to be visible on an X-ray, and more likely to cause symptoms.

Disease progression in this model is simple; neither DCIS tumours nor regional (i.e. axillary node) metastases are modelled. Separate survival functions are used for those with and without metastatic cancer. The probability of metastasis is defined as a function of tumour size, but not growth rate. Survival time with metastatic cancer

does depend on growth rate, as well as age and size. Survival time without metastatic cancer depends purely on age.

The number of life-years gained through screening is treated as a random variable, and its distribution is derived analytically from the distributions of other variables in the model. Distributions describing the disease functions are selected by fitting them to observed data using a maximum likelihood approach. The key requirement is that a basic set of model variables has been defined, and that it is complete with regard to the data. For example, fitting the model to data on the size of tumours presenting between screens requires tumour origination and growth rates to be specified, in addition to distributions on screening sensitivity and clinical presentation.

Baker uses the approach of deriving the global likelihood function for all the available data. The data used were sourced from three screening centres in north-west England, including tumour size for 761 tumours detected by screening between 1988 and 1990, tumour size and timing of presentation for 413 interval cancers presenting at the same centres, survival data for these 1174 tumours up to 1996, cancer incidence data for 1987 (before screening) and all-cause mortality for women in the region, by age. Distributions for the key model variables (in terms of functional forms and parameter values) were varied in order to maximise the global likelihood function. To be precise, the AIC was maximised. This function adjusts the likelihood for the number of parameters used in the model. It therefore penalises models that increase complexity without improving substantially the fit to the data.

The benefit of screening is in increasing life expectancy by reducing the probability of developing a metastatic cancer. The number of life-years gained through screening was treated as a random variable, and its distribution was derived analytically from the distributions of other variables in the model. The model

incorporated only the costs of each screening test undertaken for each evaluated screening programme. The fitted model suggests that screening every 3 years between the ages of 50 and 65 years (the current policy) is close to optimal if eight screens are equivalent in cost to 1 month of human life. If that cost is equivalent to 16 screens, the optimal strategy is to screen every other year between the ages of 40 and 50 years, and to screen every 3 years between the ages of 50 and 70 years.

Strengths and weaknesses of the Baker approach

A principal advantage of the Baker model is the description of variables as continuous. Many state transition models represent continuous variables such as age and tumour size in terms of banded states. Even discrete variables, such as the number of positive nodes, are usually banded to keep the number of states to manageable levels. Representing these variables as smooth functions of time is a more natural way of modelling this information. If the data gives precise values for these variables, converting them to bands wastes some of the information in the data.

Data are often available only in an aggregated form; for example, tumour size information is often reported in bands. Fitting an underlying continuous variable to this data may still have advantages as it may reduce the number of parameters that need to be fitted in the model; for example, using five bands for tumour size requires five transition rates to be defined. This could be replaced by a size variable that is a function of time with only one or two parameters. The function could be fitted to banded data by integrating over the band range; work using this approach is described below.

Baker's model characterises uncertainty parametrically, in that variables are given probability density functions with parametric functional forms. This allows a global likelihood function to be derived and minimised for all the available data. In other words, model parameters are fitted jointly using standard statistical techniques. The result is that maximum information is extracted from the data, errors around model parameters are minimised and competing parameter values can be compared in a formal rather than an *ad hoc* way. The data source for the population of the model may be restrictive for certain parameters, such as all-cause mortality, which in the model is based on a cohort of women from north-west England. A general issue for

further research is how the likelihood maximisation approach can be extended within an evidence synthesis framework.

The Baker approach has clear strengths in modelling the natural history of breast cancer and quantifying the impact of screening on disease progression. This is because many key variables cannot be directly observed, and are not naturally represented by a few discrete states. However, the effectiveness of screening is conditional on the treatment options available and as new interventions are developed, the impact of detecting tumours early will change. Screening models should incorporate a fully developed treatment model, including all prognostic factors that affect treatment options and success rates. The benefit over state transition models is less clear when modelling outcomes post-detection. Here, parameters are largely observable and can be fitted directly. Also, in practice, costs and utilities attach naturally to disease states. One approach would be to develop hybrids that combine the advantages of both types of model.

The Parmigiani model

Parmigiani[51,52] modelled probability densities for the incidence of preclinical cancer analytically as continuous functions, which were then discretised to populate an individual sampling simulation model. The Parmigiani model describes the simple transition of patients from healthy to preclinical disease to clinical breast cancer to death, with death occurring from any of the three states, while incorporating additional detail at the level of the individual:

- Transitions from healthy to preclinical are a function of age, that is, alternative transition probabilities are specified for different ages (the model follows individuals from birth).
- Transitions from preclinical to clinical are a function of age at transition to preclinical, that is, alternative transition probabilities are specified for different time periods given the age at which preclinical disease developed.
- Transitions from clinical to death are a function of age at transition to preclinical, time spent in the preclinical state and the set of prognostic factors at diagnosis. The set of prognostic factors is defined as a function of age at transition to preclinical and time spent in the preclinical state, that is, each individual samples a set of prognostic factors based on time in the previous states.

The following sections describe the analytic, and other, methods used to estimate different components of the individual sampling model.

Estimating transition probabilities to the preclinical state via deconvolution

Parmigiani states that observed distributions are available for the probability densities of:

- New clinical cases of cancer.
- Other-cause mortality (assumed to be the same for persons with and without preclinical breast cancer).
- The time in the preclinical state conditional on moving to the clinical state (three alternative assumptions/approaches are described for estimating this distribution).

Given these three densities, an integral equation is developed in which the distribution of transitions from healthy to preclinical (as a function of age) is the only unknown expression, and hence which can be solved to estimate the unknown expression. A process for discretising the integral equation is described in which the equation is reduced

> "to a matrix inversion problem. Because the inversion is typically ill behaved, a singular value decomposition followed by thresholding of low eigenvalues is useful. Once that is done, the distributions describing transitions from healthy to preclinical and from preclinical to death can be used to determine the distribution of transitions from healthy to death." (Parmigiani,[52] p. 204.)

Estimating the density of transition probabilities for progression from preclinical to clinical

Three published models of the sojourn time are considered that inform a scenario-based sensitivity analysis.

Exponential

Based on methods reported by Day and Walter,[70] Parmigiani highlights deficiencies in the assumption of an exponential sojourn time distribution, for example, the implausible assumption of a mode at zero, and the fast decay of the tail that does not adequately account for the slow-growing tumours. It is also noted that better screening tests that allow the earlier detection of cancer (e.g. smaller tumours are detectable) increase the length of the preclinical phase (which starts when the tumour becomes detectable), so estimates based on old screening data will be inaccurate.

Spratt

Spratt and colleagues define a log-normal distribution for the preclinical sojourn time (the choice of log-normal is based on independent evidence on the growth rate of tumours),[71] which defines the duration of the preclinical phase given the mean and standard deviation of the age at transition to preclinical (a different mean and standard deviation are specified for eight different age groups). The distribution is elaborated slightly to index the sojourn time distribution to an arbitrary age, and to "stabilise the location parameter (mean age at transition to preclinical) and make it monotone in age at transition to preclinical".

Peer

Peer and colleagues estimated the median doubling times [and 95% confidence intervals (CIs)] for tumours in three age groups using screening trial data.[72] Assuming tumours ⩾5 mm are detectable, that symptoms develop at 20 mm and exponential growth, Peer and colleagues developed a distribution of time in the preclinical state. Parmigiani fitted a predictive distribution of sojourn time using a log-normal with median and upper 95% quantile to match Peer and colleagues.

The estimated densities of transitions from preclinical to clinical were plotted against observed cancer incidence rates and show that the two approaches based on the exponential distribution provide the best approximation to the observed density.

The screening test sensitivity model

Parmigiani tests two scenarios with respect to defining sensitivity.[52] First, he assumes that sensitivity varies randomly between women with a uniform distribution between β_0 and 1. Second, he specifies a logit function where sensitivity is a smooth increasing function of both age and tumour size. The parameters for the equation are "loosely based" on published work, where the coefficients fit a sensitivity of 0.1 for a 0.5 cm tumour at age 60 years, a sensitivity of 0.9 for a 1.5-cm tumour at age 60 years, and a sensitivity of 0.64 for a 1-cm tumour at age 45 years. False-negative results are assumed to be independent of each other.

The axillary lymph node involvement model

This model describes a conditional probability distribution for the number of axillary lymph nodes that are found to involve metastases, which

is populated using empirical cell probabilities by age and tumour size.

Survival model

Survival considers four prognostic factors: tumour size at diagnosis; age at diagnosis; the number of positive lymph nodes; and whether the tumour is oestrogen receptor positive. A Cox regression model, with each of the prognostic factors as explanatory variables, based on combined clinical trials data is used to estimate survival. The use of trial data allows the estimated hazard rates to account for alternative treatment options.

Discussion

This chapter has illustrated the use of complex mathematical models as a means of evaluating the cost–utility of screening programmes. The model presented by Baker[10] solved a set of equations to describe the survival profile of a population of women in the absence of a breast cancer screening programme. The impacts of a range of screening programmes on that survival profile were tested. The model incorporated relevant parameters as continuous functions that change smoothly over time, although two aspects that reduce the applicability of this approach in its current stage of development include the:

- Feasibility of integrating detailed cost and utility information in order to provide more reliable estimates of the cost–utility of screening.
- Ability to incorporate additional detail in the description of the disease process; for example, the current model does not describe nodal status or treatment pathways postdiagnosis.

The already complex nature of the Baker model may mean that such model improvements are infeasible.

There are two other options for using complex mathematical models to inform cost–utility analyses of screening programmes. First, models such as that developed by Baker[10] may be used to estimate the age- and stage-specific distribution of detected cases of disease, which may than be further evaluated in a state transition model that more easily incorporates the estimation of costs and effects during the treatment phase of the disease. Although no examples of this form of cost–utility model were identified during the literature review, Duffy and colleagues describe the estimation of Markov chain models that describes progression from no disease to preclinical disease to clinical disease.[73] These models are fitted to data collected from screening programmes to estimate instantaneous transition rates between the included states simultaneously. The transition rates can be converted to transition probabilities over defined periods by solving a potentially complex set of Kolmogorov equations.[74] A five-state model is described that differentiates between node-negative and node-positive tumours, although further disaggregation of tumour states is possible. Duffy and colleagues state that "the algebraic and computing complexity render it impractical to obtain estimates for numbers of states in excess of 10 without imposing substantial further assumptions" (p. 44). An application of this work relating to the progression of colorectal tumours is presented in Chapter 4.

Second, as demonstrated by Parmigiani,[51,52] complex mathematical functions may be used to estimate separate components of a disease's natural history, which may then be used as inputs to state transition models (probably individual sampling simulation models). A large body of work already exists around the use of complex mathematical techniques to estimate different components of the natural history of cancer. The scope of the current review did not include an assessment of these modelling approaches, although work in this area published by staff at the Medical Research Council Biostatistics Unit provides a useful introduction to the types of models that have been developed, and how these models could inform parameters within a state transition screening model.

Early work concentrated on the joint estimation of disease sojourn time (the duration of the preclinical screen-detectable period) and the sensitivity of screening tests, primarily in the area of breast cancer screening.[75] Prevost and colleagues applied alternative models and estimation approaches for combined valuations of instantaneous transition rates from preclinical to clinical disease and test sensitivity for CRC.[76] Prevost and colleagues report a reasonable fit to the observed data, although they identify areas in which adaptations to the models may improve the goodness-of-fit. Sojourn time (or an aggregate estimate of the rate of clinical presentation) is unlikely to be specified directly as a screening model input parameter, although estimates of sojourn time in combination with estimates of test sensitivity may be used as additional parameters to calibrate a screening model. Gyrd-Hansen and colleagues[24,25] used earlier methods developed by

Day and Walter[70] to obtain joint estimates of test sensitivity and sojourn time for CRC, although these parameters were applied directly in a simplified natural history model (see the section 'Colorectal cancer', p. 7).

Madan developed a simple analytic model of breast cancer that provides an illustration of how established statistical methodology may be used as part of an iterative process where analytic modelling is initially used to explore and structure the problem, not to provide immediate solutions.[77] The output of interest was the distribution of the sojourn time, which was derived from distributions for the growth rate and point of presentation of the tumour. These input functions were fitted, using a similar process of likelihood maximisation to Baker, to data from the UK Breast Cancer Screening Programme describing the results of 1.35 million mammographies carried out in 2002–3. The data were banded both by age and by size of tumour detected. Nevertheless, an analytic natural history model was fitted using various functional forms for the key variables. It was shown that the distribution of tumour size for screen-detected cancers could not be reconciled with an assumption of exponential tumour growth (as used by Baker[10] and Parmigiani,[51,52] amongst others). Instead, the data were consistent with a slower growth pattern, implying a longer sojourn time. Comparing prevalence and incidence screens for women of the same age supported the idea that the sojourn time may be longer than implied by a number of models in the literature, with an approximately log-normal distribution.

Conclusions

The main conclusion drawn from this chapter is that complex mathematical modelling of the natural history of diseases has considerable potential to inform economic analyses of screening programmes, but they are underdeveloped with respect to cost–utility analyses. Further research is needed to explore the potential for full analytic models to describe additional complexities in the disease process and to provide the full range of outputs required for the cost–utility analysis of screening programmes. Further applications of partial analytic models would also be of significant interest.

Chapter 4

The analysts' perspective: a case study assessment of modelling screening for colorectal cancer

Introduction

This chapter presents a case study of a screening model for CRC, which was developed as part of a secondary cost-effectiveness analysis of screening for CRC over a 6-month period. The focus of the case study is on data sources and estimation procedures for model input parameters.

The question addressed by the case study evaluation is 'what is the likely impact of introducing various alternative CRC screening programmes in terms of incidence of cancer and mortality resulting from cancer, and in terms of costs and cost-effectiveness, for the typical population of England?' The objective is the evaluation of a range of screening programmes using FOBT and FSIG, individually and in combination, for a range of age groups.

The following section defines the model structure for the case study evaluation. The subsequent sections describe available evidence around alternative categories of input parameters, previous approaches to estimating parameter values and how the parameters were handled in the case study evaluation. These sections highlight a number of parameters that are either unobservable or for which there is great uncertainty, and the third main section describes the process of calibration for screening models. The discussion section describes limitations of the case study approach, and potential alternative approaches to various aspects of the reported evaluation of screening for CRC.

Model structure

A cohort Markov model described pathways from normal colonic epithelium to adenomatous polyp to malignant carcinoma in the general population of England. Screening interventions (and subsequent adenoma surveillance for high-risk individuals) overlaid the natural history model to estimate costs and health outcomes accrued beyond the age of 50 years over a 50-year time horizon, at which point almost all of the original cohort reached the absorption state 'death'.

Natural history model

Figure 4 describes the possible transitions allowed during each annual Markov cycle. Adenomas were classified as either low-risk or high-risk, whilst cancer states were modelled according to the Turnbull modification[78] of Dukes' staging,[79] in which metastatic disease is classified as Stage D. Lesions may develop in either the distal and proximal colon to account for the reach of FSIG. Only single lesions were described in affected individuals, although the impact of sentinel distal lesions that lead to the identification of proximal lesions by FSIG was modelled.

Due to the lack of direct evidence concerning the rate at which apparently *de novo* cancers develop, the model assumed that all cancers arise from pre-existing adenomas. Individuals may present clinically from any adenoma or cancer state.

Screening intervention model

A screening intervention model was superimposed on the CRC natural history model, which described the detection and removal of polyps, and the detection and treatment of CRC. A simplified schematic of the screening model is shown in *Figure 5*.

The impact of the different screening tests, follow-up colonoscopy and treatment of polyps and cancers were modelled by redistributing the model cohort across the health states at the point of screening. Individuals with a detected low-risk adenoma were assumed to undergo polypectomy and subsequently moved to a 'low-risk post-polypectomy' health state. Individuals identified with high-risk adenomas moved to a 'high-risk post-polypectomy' health state. Individuals remaining polyp-free for 6 years re-entered the natural history model in the normal epithelium state. Tunnel states were used to model a higher probability of recurrence in the first year after polypectomy compared to subsequent years.
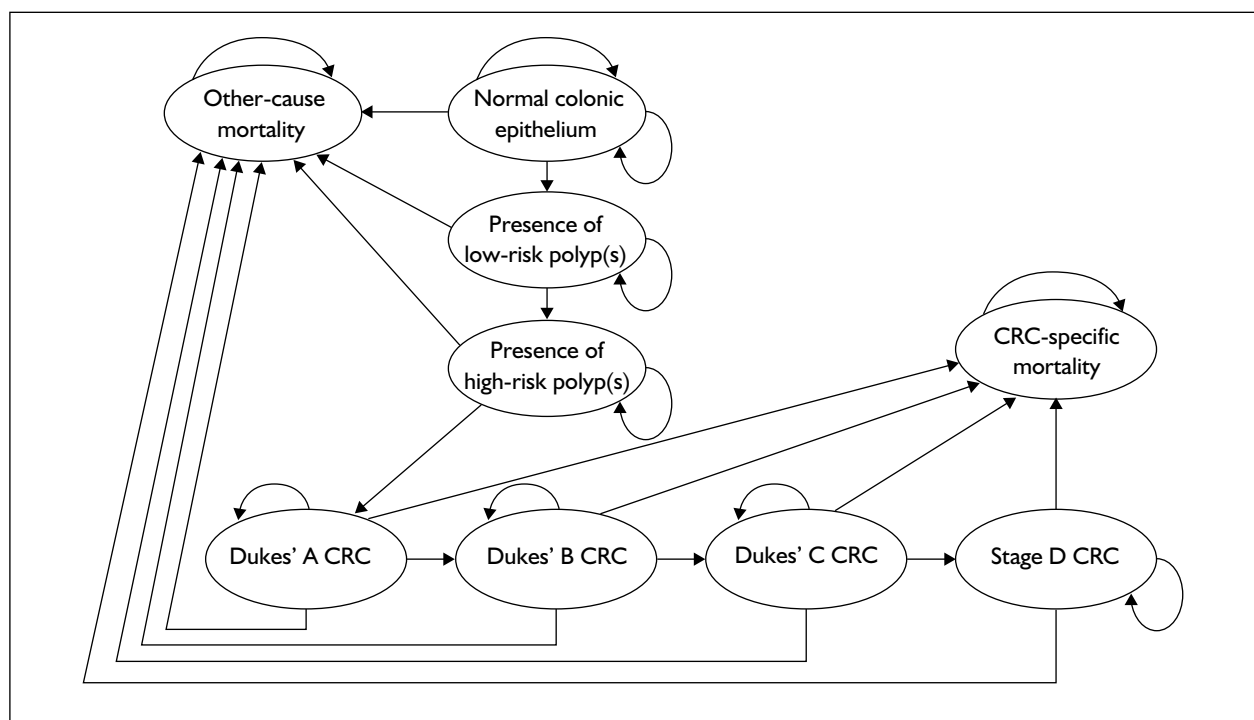
FIGURE 4 *Natural history colorectal cancer model. Separate states were modelled for each polyp and cancer state describing lesions located in the distal and proximal colon.*
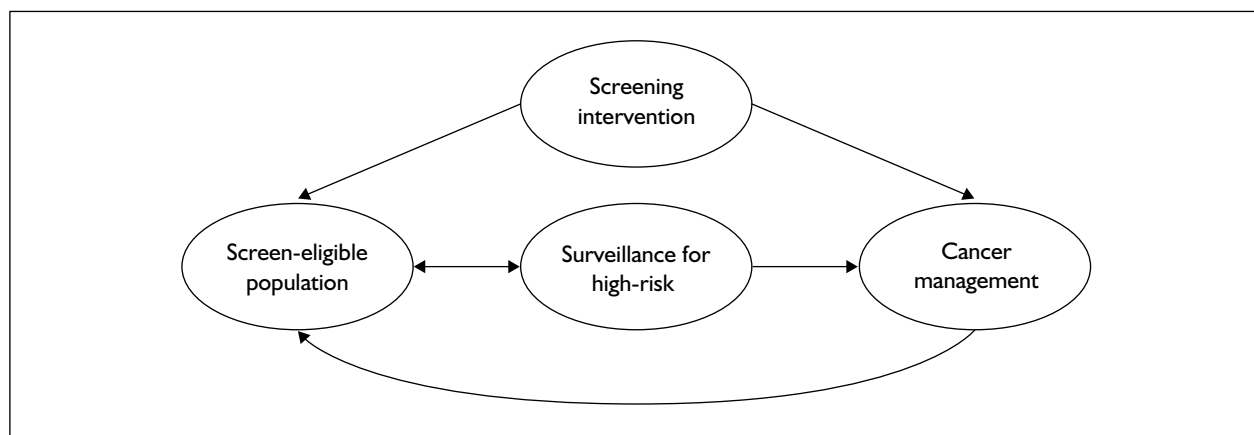
FIGURE 5 *Colorectal cancer screening model*

Individuals in whom CRC was detected by FSIG, follow-up colonoscopy or surveillance colonoscopy entered one of four 'screen-detected clinical management' health states depending on the stage of the disease at the point of detection. The model assumed no further surveillance of adenomas is undertaken beyond 80 years of age.

Assessment of research evidence

The following sections assess the research evidence for the range of categories of input parameters included in a CRC screening model. The general approach to populating the model involved the direct specification of probability distributions for parameters for which good-quality, consistent evidence was available. Feasible ranges for parameters for which evidence was either limited or inconsistent were derived from the literature and subsequently calibrated against published estimates of CRC incidence and mortality. Wide ranges were specified for parameters for which no evidence was available, which were also calibrated against published outcome data. General methods for calibrating unobserved parameters are

discussed in the section 'Progression through undiagnosed cancer states and symptomatic presentation' (p. 31).

Incidence of adenomatous polyps

Several studies of individuals undergoing surveillance colonoscopy have been published, although such studies are likely to overestimate the prevalence of colorectal adenomas as study subjects are typically at higher risk of CRC than the general population. Autopsy studies present an alternative and potentially more representative source of age-specific incidence rates, although the available studies are old. Also, as the entire large bowel is removed prior to examination, the identification of adenomas is not dependent on the sensitivity of endoscopy. Autopsy studies typically report information on the prevalence of hyperplastic and adenomatous polyps for different age groups from which crude polyp incidence rates may be estimated.

Sonnenberg and colleagues,[22,23] Vijan and colleagues[28] and Ladabaum and colleagues[26] used autopsy studies to estimate age-specific incidence of polyps, although no estimation methods are described and it is not clear how the incidence of hyperplastic polyps is handled. Wagner and colleagues state that the prevalence of polyps can be estimated from either autopsy studies or clinical trials, though their estimate appears to be based solely on autopsy records.[30] Frazier states that age- and sex-specific prevalence rates of adenomatous polyps were estimated using a weighted logistic regression analysis of six autopsy studies.[29] The OTA[31] model differentiates between the dwelling time for FOBT and other screening tests that rely upon direct visualisation of the tumour, that is, FOBT can detect smaller tumours, which require alternative estimates of polyp prevalence and incidence. Prevalence and incidence rates for polyps appear to be subjectively based on autopsy and colonoscopy studies (primarily the latter). It is not clear how the split between adenomatous and hyperplastic polyps is defined, although it is likely that this parameter is fitted to cancer incidence.

Other studies fit parameter values for polyp incidence to observed output parameters. Shimbo[27] used observed prevalence of polyps and cancers to estimate the incidence of hyperplastic and adenomatous polyps, which appears to be fitted alongside assumptions over the proportion of cancers that originate as polyps. The preclinical incidence of progressive polyps in the MISCAN–COLON model[35,36] was fitted to match CRC incidence by age, stage and localisation observed in the US prior to the advent of screening. Incidence rates for non-progressive polyps were estimated to match the observed age-specific prevalence of all polyps in a screening study.

Ness and colleagues[38] state initially that their model was populated to fit available data describing age-specific CRC incidence, CRC-related mortality and adenoma prevalence and size distribution. Eight sources are described for adenoma prevalence, although it is not clear how these data sources were combined. Adenoma incidence was calibrated to these data as a time-varying piecewise exponential function that increases with age. A single risk index for adenomatous and hyperplastic polyps was specified, which adjusted the incidence rate of new polyps for each individual. The index was fitted to match the distribution of adenoma frequency found in an autopsy study. The site distribution of polyps was assumed to be the same as that observed in the USA prior to the advent of screening.

For the case study evaluation, six autopsy studies were identified (three US studies,[80–82] two Norwegian[83,84] and one English[85]). Due to the limited reporting of the distributions of subjects within each age group, the prevalence within each age group is assumed to occur at the class mid-point. The comparison of results from autopsy studies (*Figure 6*) shows considerable differences in adenoma prevalence by age between the studies. Much of this variation is likely to be due to the limited sample sizes, particularly the small number of individuals within each age group, and geographical factors. Several of the studies are old, and may not reflect the prevalence of adenomas within the current general population due to birth cohort effects. Many of the studies suggest a reduction in the incidence of polyps within the older age groups, suggesting that the incidence of adenomas begins to plateau around the age of 65–75 years.

Given the limited information reported by the autopsy studies and the inconsistencies in the results, it is difficult to estimate the number of preclinical cancers at age 50 years. Thus, rather than estimating the prevalence of adenomas at age 50 years, the case study model begins at age 30 years, assuming that no individuals have either cancer or polyps. This assumption allows the model to 'build up' the prevalence of adenomas and preclinical cancers to the point at which
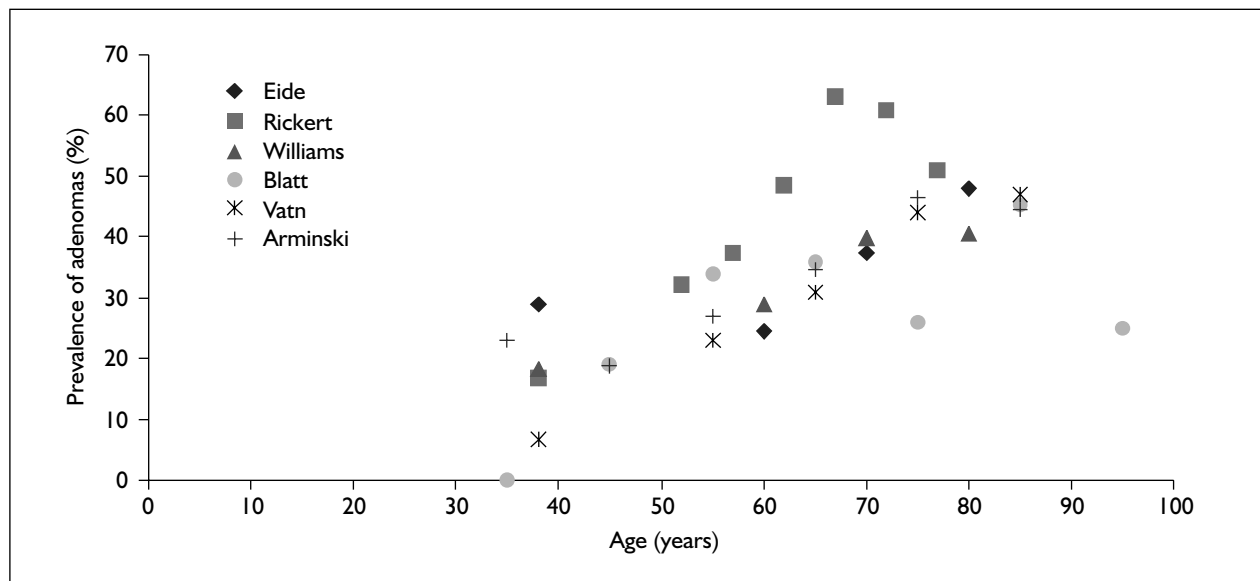
FIGURE 6 *Estimated prevalence in autopsy studies*

screening programmes are introduced. In estimating the incidence of adenomas, greatest weight was given to the study reported by Williams and colleagues[85] as this is the only study of an English population. The model assumes an annual incidence rate of around 1.5%, although this estimate was varied within the calibration process. From the age of 60 years onwards, the model assumes a proportionate reduction in the polyp incidence rate of 1.5% per year to reflect a general plateau trend apparent in *Figure 6*.

The autopsy studies provide limited information on adenoma location within each of the age groups. Given the assumption that all cancers arise from pre-existing adenomas, together with the assumption that progression rates in the proximal and distal colon are equivalent (again due to a paucity of direct evidence), the proportion of adenomas located in the distal and proximal colon was estimated by calibrating the model against published location-specific colorectal cancer incidence data.[86]

Polyp growth and transition to cancer rates

Frazier[29] based estimates of mean annual transition rates of low-risk to high-risk polyps, and from high-risk polyps to early cancer on studies of polyps left *in situ*. Wagner and colleagues state that a constant duration of 6 years was assumed for all polyps destined to become cancers,[30] which is referenced to "isolated reports of cases in which a patient refused treatment for adenoma". The OTA model report cites evidence of variation in growth rates,[31]

but also assumed a constant duration informed by subjective interpretations of observational studies of persons refusing treatment and a post-polypectomy surveillance study.

Khandker and colleagues[33] estimated a probability distribution describing the rate at which large polyps progress to cancer as predicted by a simple regression model reported by Whynes and colleagues,[87] which was based on 20-year follow-up data of unresected polyps.[88] The rate for small polyps was assumed to be one-tenth of the rate for large polyps. Vijan and colleagues[28] fitted the rate of conversion of polyps to cancer to observed cancer incidence rates, given the proportion of cancers that originate from polyps and a 10-year duration for polyps to transform from benign to malignant (based on expert opinion and indirect evidence showing the protective effect of screening). Ladabaum and colleagues[26] imply that transition probabilities between the normal, polyp and cancer states were fitted to observed data describing polyp and cancer rates.

An expert panel was assembled to inform the MISCAN model, who defined a mean sojourn time of 20 years between the onset of an adenomatous polyp and the clinical diagnosis of cancer, while the mean duration of an adenomatous polyp is stated to be 16.4 years.[35,36] The proportion of cancers arising directly from the 6–9- and the ≥10-mm polyp stages, and the proportion of hyperplastic polyps that remain in the 6–9-mm group, or progress to the ≥10-mm stage, were also estimated by the expert group.

Ness and colleagues[38] reference studies informing assumptions of mean sojourn times of 26 and 52 years for the fast- and slow-growing adenomas, although no details of the type of data are reported.

For the case study evaluation, three studies reporting growth rates of polyps left *in situ* were identified.[89–91] Knoernschild reported the largest number of polyps over the longest follow-up period, and so these data were used to inform progression rates between low- and high-risk polyps.[91] A total of 213 of 257 patients underwent a repeat sigmoidoscopy between 3 and 5 years, which showed that approximately 1% of small low-risk polyps progressed to high-risk polyps each year. As with the estimates of polyp incidence, this estimate was allowed to vary during the model calibration process.

The rate of malignant transformation from adenoma to carcinoma was based on a retrospective review of records over a 6-year period in 226 patients with diagnosed colonic polyps ≥10 mm in diameter,[88] in whom periodic radiographic colonic examination was elected over surgical excision. In most cases the polyp was left *in situ* due to poor medical condition which precluded resection. About 2.5, 8 and 24% of the remaining study cohort had developed invasive colorectal cancer at 5, 10 and 20 years, respectively. Only 14 of 226 patients were followed up to 20 years. Loss to follow-up was primarily attributable to the subsequent excision of non-malignant polyps due to growth, which likely led to an underestimate of invasive cancer rates, that is, the highest risk polyps were removed prior to transition to cancer. However, presented rates implicitly assume that polyp incidence was at the time of entry to the study, whereas some of the individuals may have had adenomas for a long period before entering the study. This assumption leads to an overestimate of polyp transition rates. At best, Stryker and colleagues[88] demonstrate that the adenoma–carcinoma sequence is generally slow, and may take 10 years or longer.

Exponential and Weibull survival curves were fitted to the published data points using the least-squares approach in order to extrapolate the risk of developing cancer beyond the follow-up period. *Figure 7* shows that the Weibull and exponential survival curves are similar, suggesting a constant risk of invasive cancer of roughly 1.5% per year. Due to the uncertainty around this estimate, the transition probability from high-risk adenoma to colorectal cancer was included in the model calibration process.

Progression through undiagnosed cancer states and symptomatic presentation

There exists no direct empirical evidence describing the rate at which patients progress through preclinical cancer states prior to
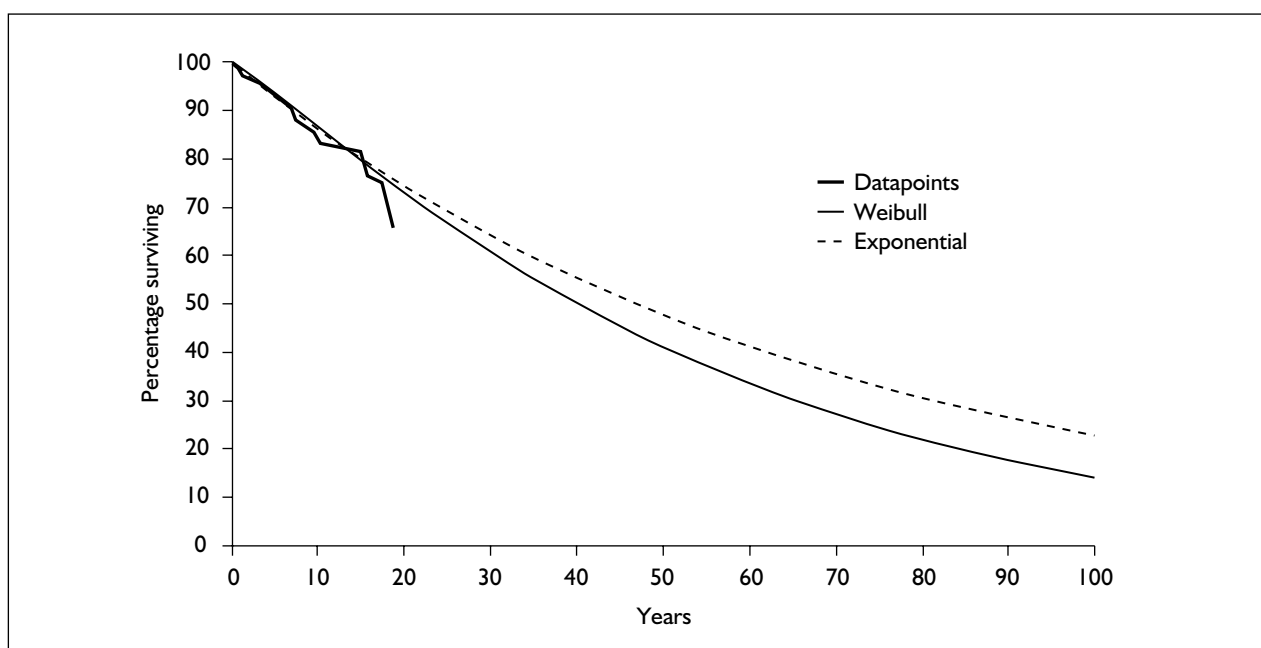

FIGURE 7 *Fitted survival curve to estimate progression of adenomas to invasive cancer*

diagnosis, or the probability that an individual with preclinical disease will present symptomatically within a given period. These parameters were estimated as part of a full model calibration in most of the reviewed screening models, although Loeve and colleagues undertook a partial fitting process to estimate sojourn times in cancer states (I–IV) based on the ratio between the state-specific detection rate at first screening in FOBT trials and the background incidence, accounting for the sensitivity of FOBT for all cancer stages.[35,36]

In the case study evaluation, preclinical progression rates were fitted during the model calibration process without specifying clinically plausible ranges. Clinically plausible ranges describing the probability of symptomatic presentation with CRC were based on an existing screening modelling study,[29] and were allowed to vary stochastically during the model calibration process.

An alternative partial calibration approach was also applied, which built on the use of Markov chain models that describe progression from no disease to preclinical disease to clinical disease, as described in the section 'Discussion' (p. 24).[73,92] The states and transitions represented in the CRC Markov chain natural history model are presented in *Table 2*. Duffy and colleagues reported a five-state model describing breast cancer natural history;[73] the CRC adaptation was a nine-state continuous time Markov model. The prototype model estimates the probability that in a very small period an individual can only move from their current state to the next worse health state or present symptomatically (represented as λ_j). Markov chain Monte Carlo (MCMC) simulation was used to fit the transition rates λ_1–λ_8, and screening test sensitivity rates to event probabilities from screening trials and national incidence data. The annual transition probabilities presented in *Table 3* were estimated by solving Chapman–Kolmogorov equations.

These preliminary results overestimated incidence for low ages and underestimated for high ages, as demonstrated in *Figure 8*.

This developmental work demonstrated the feasibility of this type of approach, although further work is required to develop an evidence synthesis model of the natural history of CRC that can be incorporated in a cost–utility analysis of screening programmes for CRC. Applied developments may include amendments to the model structure, the estimation of age-specific transition rates and incorporating additional data sources. Methodological developments include assessing convergences to supporting distributions on the MCMC sampler, the use of more formal model comparison techniques and undertaking probabilistic sensitivity analysis of health economic outcomes incorporating correlation in the uncertainty in natural history parameters and test characteristics.

Correlations between the distal and proximal colon

The case study Markov model did not explicitly model the incidence and development of separate lesions within individuals. To overcome this limitation, the model described a proportion of proximal lesions as having sentinel distal lesions,[93,94] which lead to the detection of proximal lesions during further investigation and workup following FSIG. This input parameter is informed by Dinning and colleagues,[93] who undertook a retrospective study of 634 colorectal adenomas and found no 'sentinel' distal lesion in 72% of cases of proximal cancer.

TABLE 2 *States and transitions represented in the CRC Markov chain natural history model*

State	Possible transitions to
Well	Low-risk polyp (λ_1)
Low-risk polyp	High-risk polyp (λ_2)
High-risk polyp	Dukes' type A preclinical (λ_3)
Dukes' type A preclinical	Dukes' type A clinical (λ_4); Dukes' type B or C preclinical (λ_5)
Dukes' type A clinical	–
Dukes' type B or C preclinical	Dukes' type B or C clinical (λ_6); Dukes' type D preclinical (λ_7)
Dukes' type B or C clinical	–
Dukes' type D preclinical	Dukes' type D clinical (λ_8)
Dukes' type D clinical	–

TABLE 3 *Estimated 1-year transition probabilities from prototype Bayesian synthesis model*

|	Normal	Low-risk	High-risk	Preclinical A	Preclinical B&C	Preclinical D	Clinical A	Clinical B&C	Clinical D	Dead
Normal	0.9935	0.002407	6.88×10^{-5}	3.31×10^{-7}	1.69×10^{-7}	1.03×10^{-7}	8.41×10^{-8}	2.57×10^{-7}	1.64×10^{-7}	0.004
Low-risk	0	0.9421	0.05261	3.10×10^{-4}	1.67×10^{-4}	1.23×10^{-4}	1.11×10^{-4}	3.59×10^{-4}	2.58×10^{-4}	0.004
High-risk	0	0	0.9498	0.005969	0.003844	0.003335	0.004353	0.01566	0.0131	0.004
Preclinical A	0	0	0	2.58×10^{-4}	0.001017	0.007843	0.1047	0.4262	0.4422	0.004
Preclinical B&C	0	0	0	0	7.41×10^{-6}	0.002028	0	0.4517	0.4754	0.064
Preclinical D	0	0	0	0	0	5.63×10^{-4}	0	0	0.5754	0.424
Clinical A	0	0	0	0	0	0	0.996	0	0	0.004
Clinical B&C	0	0	0	0	0	0	0	0.936	0	0.064
Clinical D	0	0	0	0	0	0	0	0	0.576	0.424
Dead	0	0	0	0	0	0	0	0	0	1
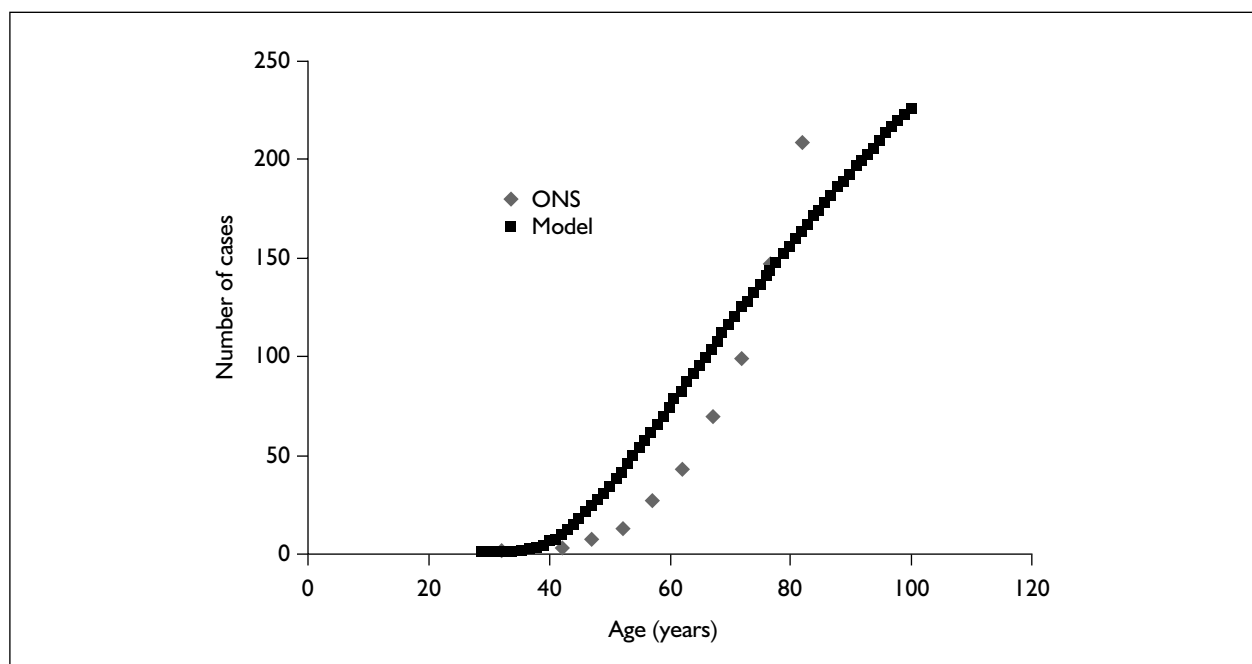
FIGURE 8 *Cancer incidence estimated by the colorectal cancer Markov chain natural history model. ONS, Office for National Statistics.*

Recurrence rates of adenomas following polypectomy

Goldie and Kuntz[95] highlight a potential error relating to the application of higher transition probabilities to the cancerous states for individuals with a previously removed polyp, which may result in counterintuitive results. Defining a progression ratio as

$$\text{progression ratio} = \frac{Pr(\text{progression to cancer} \mid \text{previous polyp})}{Pr(\text{progression to cancer} \mid \text{no previous polyp})}$$

As the progression ratio increases, the aggregate risk of cancer for individuals with precancerous lesions would be less if they remained undetected than if they were all detected and a proportion of that population experienced new lesions and the higher risk of progression. Goldie and Kuntz conclude that unless there is specific information describing differential progression rates in the two populations, it is advisable to assume constant transition probabilities across all individuals with precancerous lesions.

In the case of colorectal cancer, empirical evidence strongly suggests that individuals with a history of adenomatous polyps are more likely to develop subsequent polyps than individuals without polyps.[96–98] Estimates of recurrence rates for individuals with low-risk and high-risk adenomas following polypectomy were derived from a comparative study of the effectiveness of colonoscopy performed after the colonoscopic removal of adenomatous polyps.[98] From this study, adenomas larger than 10 mm are defined as high-risk and other polyps are classified as low-risk. The risk of developing a new low-risk adenoma given a history of prior low-risk adenomas was estimated to be 18% in the first year following excision and 5% each year thereafter. For individuals with previously excised high-risk polyps, the risk of developing a new low-risk polyp was estimated at 25% in the first year after excision and 6% per year thereafter.

Test characteristics

The majority of the reviewed screening models referenced sensitivity and specificity rates to various sources but provided no explanation of the estimation methods and assumptions. Gyrd-Hansen and colleagues[24,25] used methods of maximum likelihood estimation to estimate test sensitivity and sojourn time,[70] which were applied to a simplified natural history model. More complex methods, as reported in the section 'Progression through undiagnosed cancer states and symptomatic presentation' (p. 31), would be required to apply these methods to the case study model. Shimbo and colleagues estimated binomial confidence intervals[99] using the sum of total cases, although details of the data included are not provided.[27]

The following sections describe identified data sources for the estimation of sensitivity and specificity for FOBT, FSIG and colonoscopy.

Faecal occult blood testing

Three general methods for the estimation of screening test sensitivity are:

- 'screening' patients with known CRC
- 'screening' before and independently from diagnostic evaluation
- diagnostic work-up of test-positive individuals and the follow-up of test-negative individuals to identify missed cases.

Windeler and Kobberling report a mean sensitivity rate of 76% of Hemoccult FOBT for cancer from a meta-analysis of 10 studies of patients with known CRC, while meta-analysis of 12 studies reporting tests undertaken prior to diagnostic evaluation estimates a sensitivity rate of 65%.[100] Sensitivity rates of 10 and 20% are reported for all adenomas and adenomas larger than 1 cm, respectively.

Two studies reported by Allison and colleagues estimated sensitivity on the selected work-up of positive screen results and the follow-up of test-negative individuals.[101,102] The combined studies included 23,292 individuals screened using Hemoccult II with follow-up to 2 years. The combined results suggest that the sensitivity of FOBT in detecting CRC is 40.51%, which informed the case study evaluation. The combined sensitivity for adenomas was estimated to be approximately 29.2%, although as very few polyps present symptomatically this is likely to be a considerable overestimate. In the absence of more robust methods for evaluating the true sensitivity of the test, the case study model assumed a sensitivity in detecting adenomas of 10%, which is in line with the meta-analysis[100] and several other modelling studies.[29,33,103]

Alternative approaches include combining the observed screen-detection rate of adenomas with the estimated prevalence in the screening model. From the section 'Incidence of adenomatous polyps' (p. 29), the prevalence of adenomas at age 55 years is estimated to be around 25%, and Allison and colleagues report a detection rate of 0.3%. Test sensitivity for adenomas is estimated as $0.3/25 = 1.3\%$.

Work-up bias is introduced when subjects with positive or negative diagnostic test results are selectively referred to receive verification by the validating criterion. If it can be assumed that the patients undergoing the reference test are randomly sampled from their respective populations of positive and negative screen results, then the Bayes theorem equation can be used to estimate sensitivity and specificity:

$$P(R+/D+) = \frac{P(R+) \times (P(D+/R+,V+)}{\begin{array}{c} P(R+) \times P(D+/R+,V+) + \\ P(R-) \times P(D+/R-,V+) \end{array}}$$

where $P(R+)$ = probability of a positive screen result in the full population, $P(D+/R+,V+)$ = positive predictive value in the biopsy population, $P(R-)$ = probability of a negative screen result in the full population and $P(D+/R-,V+)$ = 1 minus negative predictive value (NPV) in the biopsy population.

In some cases, samples are not randomly selected; for example, in split sample studies, tests that are negative for one screening test may be positive for the other screening test and hence sent for biopsy, such that the negative screens sent for biopsy are not representative. The NPV will tend to be underestimated, which lowers the estimated sensitivity. In such cases, an adaptation to the work-up bias correction method assumes a constant NPV for studies with suspected diagnostic work-up bias. A range of estimated NPV values may be used to assess the robustness of the resulting estimates of sensitivity and specificity.

Colonoscopy and flexible sigmoidoscopy sensitivity

The sensitivity of colonoscopy and FSIG is dependent on the adequacy of the bowel preparation and the competence and experience of the endoscopist. Recent evidence suggests that detection rates may vary considerably between different centres.[104] Two studies assessed colonoscopic miss rates of adenomas using back-to-back colonoscopies undertaken on the same day.[105,106] A third study undertook a retrospective analysis of colonoscopic miss rates in individuals with a new diagnosis of colorectal cancer over a 3-year follow-up period.[107] This study suggests that the miss rate may be dependent on the location of the adenoma or cancer, with higher miss rates in the proximal colon. The data from these studies were combined to estimate separate miss rates for small and large adenomas and for cancer. Higher miss rates were also estimated for large polyps and cancers located in the proximal colon. Due to limited evidence on the sensitivity and specificity of FSIG in usual clinical practice, with current evidence restricted only to case–control studies,[108] miss rates for FSIG were

assumed to be identical with those for colonoscopy within the distal colon.

Inadequate rates for endoscopy
In a small number of cases, endoscopic examination may need to be repeated due to inadequate bowel preparation. The UK FSIG trial[109] reported that approximately 5% of tests had to be repeated due to inadequate bowel preparation. In cases whereby colonoscopy is incomplete due to inadequate bowel preparation, it is likely that barium enema would be used instead (Atkin W, Colorectal Cancer Unit, St Mark's Hospital, Harrow: personal communication, 2004). Recent guidance on the use of endoscopy in England and Wales estimated that approximately 10% of colonoscopies are not completed due to inadequate bowel preparation.

Harm caused by screening
Similar estimates of the probability of perforation following FSIG of 0.002% were reported by two separate studies, involving a combined total of almost 150,000 tests.[109,110] The UK FSIG trial also reported four perforations in 2377 colonoscopies (with polypectomy).[109] The model assumed that the risk of perforation associated with colonoscopy with polypectomy is 0.168%. Based on expert opinion, the probability of perforation for colonoscopy without polypectomy was assumed to be half of the risk of colonoscopy with polypectomy.

Probabilities of 0.0295 and 0.439% for hospital admissions due to bleeding following FSIG and colonoscopy, respectively, were also informed by the UK FSIG screening trial.[109]

Mortality
The model incorporates three causes of mortality: death due to causes other than CRC, death due to CRC and death due to perforation of the bowel. Age-specific probabilities of dying from causes other than CRC were estimated using standard life expectancy tables obtained from the Government Actuary's Department,[111] subtracting the age- and sex-specific risk of death due to CRC using

mortality estimates obtained from the Office for National Statistics (ONS).

Probabilities of death due to perforation of the large bowel were derived from a US study.[97] Of the 77 perforations due to colonoscopy, four patients died within 14 days (5.19%). Of the 31 perforations following FSIG, two patients died within 14 days (6.45%). Due to the limited number of subjects included in the analysis, a common probability of dying following bowel perforation of 5.82% was applied.

Most studies used stage-specific mortality rates, primarily derived from the US National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) database. Variations include Khandker and colleagues,[33] who state that 5-year survival rates were based on stage at diagnosis and years with cancer, although it is unclear how the impact of the duration of cancer was implemented. Wagner and colleagues compared age-specific 5-year survival probabilities for late-stage cancer with the expected age-specific survival for the normal population, to estimate the gain in life-years from the prevention of a single case of late-stage CRC.[30] The surviving proportion at 5 years was assumed to be cured, and the dead proportion was assumed to die at 3 years.

UK data from an audit study undertaken in the Wessex region between 1991 and 1995 informed the case study evaluation. This study estimated crude CRC mortality rates from patient-level data describing 5-year follow-up of 5173 patients.[112] Survival analyses for deaths due to CRC and to other causes were undertaken by age and stage at diagnosis using data from 4872 patients. Estimates of mortality (averaged across all age groups) are given in *Table 4*. The estimated annual mortality rates were allowed to vary within tight ranges within the model calibration process.

Uptake rates for screening and follow-up
None of the reviewed studies linked uptake rates to cost or disease progression parameters. Some

TABLE 4 *Annual mortality estimates by stage and cause*

Cancer stage	Probability of death due to CRC (%)	Probability of death due to any other cause (%)
Dukes' A	1.25	3.76
Dukes' B	5.04	3.42
Dukes' C	13.25	4.21
Dukes' D	40.74	4.63

studies simply assumed 100% uptake,[27,33] Vijan and colleagues assumed less than 100% uptake, but assumed a constant uptake rate across programmes,[28] whereas Frazier specified separate uptake rates for different screening programmes observed in clinical trials.[29] Frazier also states that it may be important to specify uptake rates by screening round for screening programmes that involve repeat screening rounds.

Due to the timescale of the case study evaluation, uptake rates were also modelled as independent parameters. It was assumed that a proportion of individuals will never attend screening, whilst the remaining proportion has full compliance with all screening rounds. The same uptake rate of 60% was assumed for both FOBT and FSIG, based on similar compliance rates in the demonstration pilot[103] and the Nottingham RCT[113] for FOBT. The UK FSIG trial[109] included only those individuals who had expressed a prior interest in screening, so compliance rates observed within the trial are unlikely to reflect compliance rates for a national screening programme.

A 77.5% attendance rate for follow-up colonoscopy was derived from the UK demonstration pilot,[103] which is similar to that reported by Lund and colleagues.[114] The same compliance rate was assumed in the case study model for colonoscopy following a positive screening test, and as surveillance of individuals in whom high-risk adenomas are detected.

Costs

The timescale of the case study evaluation precluded a detailed cost analysis, and so resource assumptions were based primarily on the costing assumptions detailed within the FOBT CRC Pilot Evaluation Report,[103] a Health Needs Assessment report on the management of CRC[115] and input from members of the Bowel Cancer Advisory Group. The majority of the costs used were derived from the NHS Reference Cost Schedules from 2003,[116] Unit Costs of Health and Social Care[117] and estimates from other relevant studies.[118–120] The analysis includes direct costs only, which were inflated to 2004 prices using NHS Health Care Inflation Indices.[117]

The unit cost of FOBT is based on programme estimates incorporating two test kits (repeat tests and non-compliance), analysis of the slides and sending letters informing GPs of the test results (Patnick J, NHS Cancer Screening Programme: personal communication, 2004). Administration costs of £1.74 per person invited (for obtaining

patient details and sending out initial invitation letters) were estimated from an audit of the Cervical Screening Programme.[121]

Costs of undertaking colonoscopy and FSIG are available from the NHS Reference Costs based on Healthcare Resource Group (HRG) categories, which are of a similar magnitude (£165 and £142, respectively). However, comparison with published patient-level cost estimates shows significant variation between the cost estimates for FSIG.[118] Closer analysis reveals that FSIG and colonoscopy typically take place in the same endoscopy suite, with some hospitals undertaking a combination of both procedures within the same session. Typically five or six colonoscopies per day can be performed compared with 10–12 FSIGs per day. In addition, more expensive equipment is required for a colonoscopy procedure. It is therefore expected that cost of a colonoscopy would be at least twice that of an FSIG. This demonstrate the potential problems of using reference costs that are based on varying costing methodologies, and the unit cost for FSIG reported by Whynes and colleagues[118] is used in the model.

Resources associated with diagnosis of screen-detected and clinically presenting patients were informed by expert opinion. Differential costs for the histopathological analysis of colorectal adenomas and carcinomas were informed by expert opinion, reflecting the cost of repeat tests plus clinician time in presenting results at multi-disciplinary team meetings.

The lifetime costs of cancer management are applied within the model as a one-off cost dependent on the stage of the cancer at diagnosis, which include endoscopic treatment (for polyps), preoperative radiotherapy, primary radical radiotherapy, surgery, postoperative radiotherapy, postoperative chemotherapy, surveillance, recurrence and palliative care.

Data from the UK FOBT pilot demonstration study showed that approximately 23% of screen-detected Dukes' A cancers can be successfully treated using endoscopy without the need for surgical resection.[122] Staging data from the Wessex Audit[112] were used to estimate the proportion of patients who would undergo preoperative radiotherapy.

Estimates of stage-specific surgery, primary radical radiotherapy and postoperative chemotherapy rates, surveillance schedules for cancer patients, costs of treating recurrent disease and costs of

palliative care are based on the expert opinion of members of the case study steering group.

Health-related quality of life

Only three studies reporting health-related quality of life associated with CRC were identified.[123–125] Dominitz and Provenzale[125] measured utilities for a single, global description of CRC in individuals with and without CRC using the time trade-off method. This global description did not include the impact of cancer stage on health outcomes. Whynes and Neilson[123] estimated population-based utility scores for outcome states of CRC, but found little difference between cancer stages. Their valuations of different CRC health states were extremely close to valuations for perfect health (i.e. a valuation of 1.0).

Ness and colleagues[124] assessed utilities associated with stage of cancer and treatment. The study recruited 90 individuals who had previously undergone removal of a colorectal adenoma. Individuals were interviewed and were asked to assess utilities for stage-dependent outcome states using the standard gamble technique. Seven health states for CRC were presented to the study participants, although only five scenarios were presented during each interview. The scenarios described specific areas of morbidity associated with CRC and treatment such as tiredness and weakness, changes in bowel habits, sexual problems, pain, cognitive problems, social problems, and emotional problems. The results of the analysis for each of the individual CRC-specific health states are shown in *Table 5*.

In the case study model, the utility associated with all cancer- and polyp-free states, and also for individuals with undiagnosed cancer, was assumed to be equivalent to the 'no known adenomas' health state (utility score = 0.91). Mean utility scores of 0.74 for Dukes' A cancer and 0.27 for Stage D cancer were used. As only aggregate stage states were modelled, mean utility values of 0.7 for Dukes' B cancer (high score) and 0.5 for Dukes' C cancer (low score) were assumed. Utility estimates were not adjusted to account for the impact of co-morbidities associated with age.

Model calibration

The previous section described the data and methods used to populate the case study model, although a number of unobservable parameters were identified for which direct estimates could not be defined. There were also a number of parameters for which the data were so uncertain as to preclude the reasonable estimation of a mean parameter value, for example, polyp incidence rates and the transition rate of adenomas to CRC. Most of the reviewed cancer screening models calibrated unknown parameter values to national incidence rates and detection rates from RCTs of screening programmes, implying that parameter values were adjusted manually until the model 'fitted' the data. However, there may exist numerous potential model solutions which 'fit' the data, which raises important doubts over the validity of the results of these analyses.

Draisma and colleagues[126] describe a representative process of calibration for a prostate cancer screening model to prescreening cancer incidence and data from a screening trial. Alternative sets of input parameter values were run through the model and the best fitting set of parameters was defined by minimising the difference between the predicted and observed estimates for each output parameter, measured as the sum of the χ^2 quantities using an adapted version of a simplex optimisation method.[127] Church expresses concerns about this approach due to the degree of dependence between the output parameters to which the input parameters are fitted.[128] He states that although the degrees

TABLE 5 *Utility scores used to describe health-related quality of life*[124]

CRC-specific health state scenario	Utility score (95% CI)
No known adenomas	0.91
Stage I rectal or Stage I/II colon cancer treated with resection only	0.74 (0.69 to 0.78)
Stage II colon cancer treated with resection and chemotherapy without significant side-effects	0.70 (0.63 to 0.77)
Stage II colon cancer treated with resection/chemotherapy with significant side-effects	0.63 (0.56 to 0.70)
Stage II/III rectal cancer treated with resection/chemotherapy/radiation therapy	0.59 (0.54 to 0.64)
Stage II/III rectal cancer treated with resection/chemotherapy/radiation therapy/permanent colostomy	0.50 (0.44 to 0.56)
Stage IV metastatic/unresectable disease without colostomy	0.24 (0.16 to 0.32)
Stage IV metastatic/unresectable disease with colostomy	0.27 (0.18 to 0.36)

of freedom, upon which the sum of the χ^2 was estimated, were reduced to account for the marginal constrainment of some of the output observations, the adjustment on the degrees of freedom does not account for the full range of dependences between the output parameters. The impact is that more than one set of input parameters could have the same fit as the defined 'best-fitting' parameter set.

Case study calibration methods

The model parameters to be calibrated included unobservable parameters (transition probabilities between undiagnosed cancer states and probabilities of clinical presentation by cancer stage) and parameters informed by weak sources of evidence (polyp incidence and growth rates, the rate at which high-risk adenomas develop into cancer, and stage-specific CRC-specific mortality rates).

The first stage of the calibration process for the case study model involved fitting natural history parameters in the absence of screening. The following output parameters were identified:

- English CRC age and site-specific incidence rates[86]
- English CRC age-specific mortality rates[129]
- adenoma prevalence by age[80–85]
- stage distribution of symptomatic CRC at diagnosis[112]
- annual CRC-specific mortality rates by stage.[112]

The second calibration stage used unpublished data from the Nottingham FOBT trial and the UK FSIG screening trial[109] to calibrate single test sensitivities for FOBT and FSIG in detecting adenomas and cancers.

A pragmatic approach to model calibration was used within the case study. Each unknown model parameter was assigned a uniform distribution, preferably based on some evidence from the literature. About 60,000 sets of input parameter values were randomly sampled and the mean squared errors between the model predictions for each input parameter set and published stage-specific incidence and prevalence data and mortality data were recorded. A subjective threshold for the degree of 'acceptable' error was specified, whereby all solutions below the acceptability threshold gave a reasonable visual fit against the published data sources. Greater subjective weight was placed on how the model fitted against published incidence and mortality, rather than adenoma prevalence due to inconsistencies in the published evidence on adenoma prevalence. The process for identifying potential parameter sets is described in *Figure 9*.

The main disadvantage of this approach is the use of subjective acceptability thresholds, although this allowed for greater flexibility in the potential parameter sets than the specification of an arbitrary threshold, for example, accept all combinations within 10% of mean observed values.

Case study calibration results

Of the 60,000 random iterations, around 400 potential solutions were identified. *Table 6* presents

TABLE 6 *Natural history transition probabilities estimated by calibration*

Transition probability (annual)	Uniform distributions tested (%)		Ranges in included parameter sets (%)	
	Minimum	**Maximum**	**Minimum**	**Maximum**
Normal epithelium to low-risk polyp (aggregate)	0.5	2.0	1.5	1.8
Low-risk polyp to high-risk polyp	0.5	4.0	1.5	4.0
High-risk polyp to Dukes' A	1.0	6.0	1.7	6.0
Dukes' A to Dukes' B	30.0	90.0	30.1	89.9
Dukes' B to Dukes' C	30.0	90.0	40.1	90.0
Dukes' C to Dukes' D	30.0	100.0	75.0	100.0
Probability of symptomatic presentation (Dukes' A)	2.0	15.0	5.1	15.0
Probability of symptomatic presentation (Dukes' B)	10.0	35.0	10.1	34.9
Probability of symptomatic presentation (Dukes' C)	50.0	90.0	45.1	79.9
Probability of symptomatic presentation (Stage D)	50.0	90.0	70.1	100.0
Annual CRC-specific mortality rate (Dukes' A)	0.0	0.5	0.0	0.0
Annual CRC-specific mortality rate (Dukes' B)	0.5	3.0	0.5	3.0
Annual CRC-specific mortality rate (Dukes' C)	2.0	15.0	4.0	8.0
Annual CRC-specific mortality rate (Stage D)	35.0	45.0	35.0	43.0
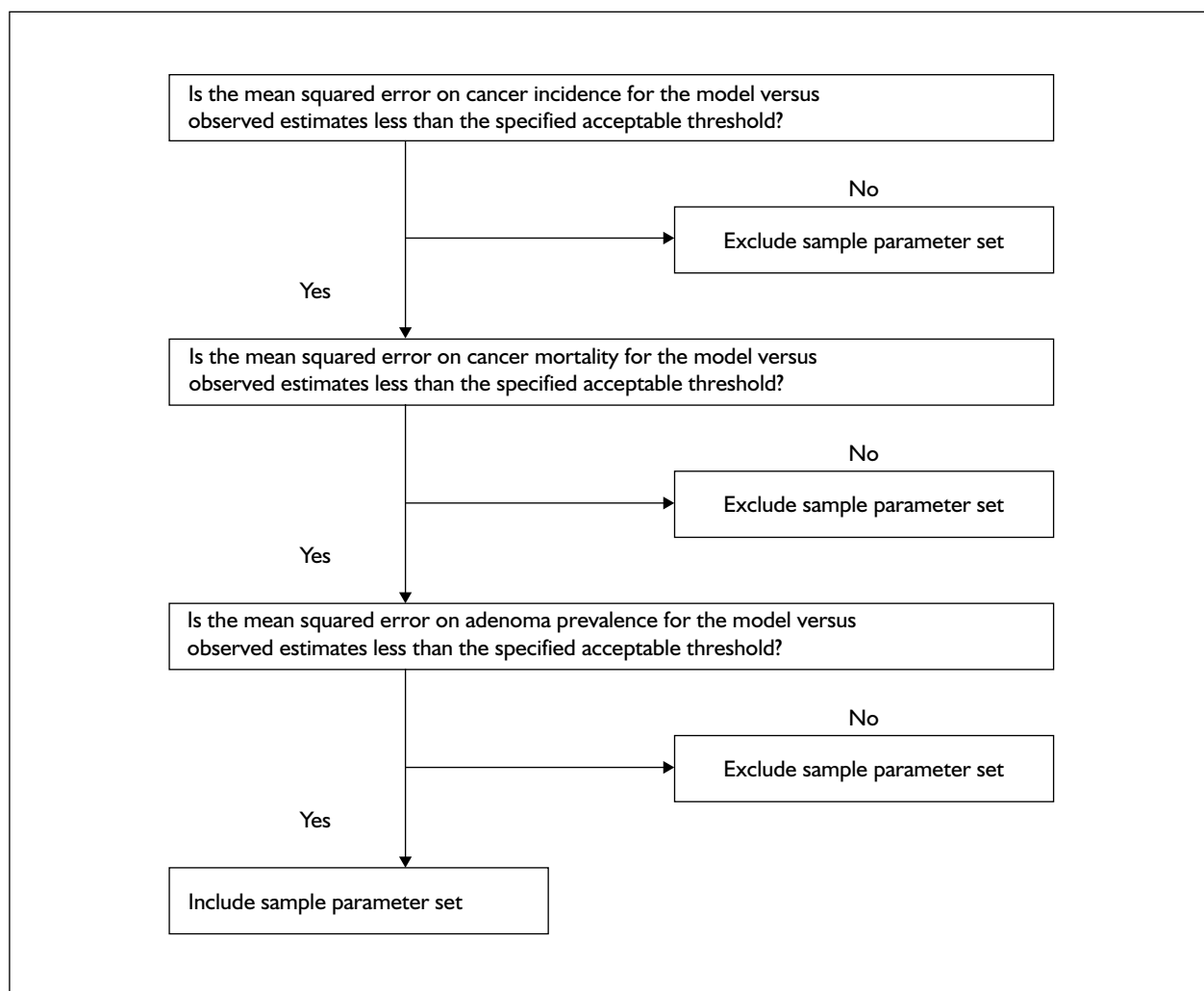
FIGURE 9 *Process for identification of potential parameter sets*

the range of values tested, and accepted, for each calibrated input parameter.

Discussion

This chapter has described the methods and data sources used to develop and populate a case study model of screening for CRC; the results of this model are not presented here, but are available in detail elsewhere.[130] The following sections discuss the advantages and limitations of the case study evaluation.

Model structure

The potential limitations of the case study model structure are partly related to the choice of modelling technique (a Markov state transition model), which was discussed in Chapter 2. First, the incidence and progression of multiple lesions within individuals were not explicitly modelled,

although a proportion of proximal cancers were assumed to be associated with 'sentinel' distal polyps. Ideally, this proportion should be an output of the model, not a parameter input. Individual sampling models describe the development of individual lesions within individuals,[35,36] which may be viewed as a more intuitive approach.

Second, transition probabilities estimated within the model are assumed to be constant (with the exception of age-specific adenoma incidence and all-cause mortality rates). Tunnel states could be used to incorporate time-varying transition probabilities in a Markov model, although the additional complexity precludes the use of tunnel states as a feasible option.

Other simplifying assumptions include that all cancers derive from pre-existing adenomas, despite some indirect evidence that a proportion

TABLE 7 *Number of cancers remaining undetected for alternative systematic false-negative rates given an individual test sensitivity of 50%*

Screening round	25% systematic false negatives	0% systematic false negatives
1	1000	1000
2	532	500
3	283	250
4	151	125
5	80	63

of CRCs arise *de novo*. Relaxation of this assumption requires no additional states, just the possibility of transiting from the normal epithelium state to the Dukes' A cancer distal or proximal states, which increases the number of unobservable parameters to be estimated via calibration.

The model does not explicitly describe individuals who would be classified as intermediate risk according to the BSG guidelines (multiple small adenomas).[131] The further disaggregation of the polyp states would introduce additional unobservable parameters, such as progression rates and differential screening test sensitivities, which it was felt would not materially affect the results of the model.

Loeve and colleagues[67] investigated the impact of systematic false negative results, under the hypothesis that a proportion of lesions will remain undetected by FOBT because they will never bleed. *Table 7* illustrates the impact of systematic false-negative results, which shows that for 1000 cancers that remain undetected after the first screening round, after four further screening rounds 17 fewer cancers would be detected if 25% of the false-negative results were systematic.

To model the existence of systematic false-negative lesions in a cohort Markov model approach would require the specification of separate sets of health states to describe the progression of lesions to which a non-systematic, and a systematic, false-negative rate could be applied, that is, a doubling of the preclinical health states. Using an individual sampling model, Loeve and colleagues[67] were able to specify the nature of a lesion as an attribute that determined the appropriate false-negative rate for each lesion.

The above adaptations would introduce greater model complexity and/or a greater number of input parameters that require fitting to observable outputs, which would increase the uncertainty

around the set of calibrated input parameters. The appropriate balance between model complexity, the practical implementation of a model and the uncertainty around the calibration process is subject to the interpretation of the research team. There are no general rules as the impact of simplifying assumptions will vary between disease areas. Decisions regarding the choice of model structure should incorporate the views of experts in the disease area being modelled and the previous experience of the analyst, and justifications for the choices made should be explicitly stated.

Model population

The following sections describe the issues arising from the estimation of key input parameters for which direct evidence were limited, or for which alternative methods of parameter estimation are available.

Polyp incidence and prevalence

Some studies imply that age-specific incidence rates can be estimated directly from observed prevalence rates, for example, if polyp prevalence at age 50 years is estimated to be 10% and prevalence at age 55 years is 20%, then the annual incidence rate between the ages of 50 and 55 years is estimated as

$$\text{Annual incidence rate} = \frac{1 - \exp\{\ln[1 - (0.1/0.9)]/5\}}{= 2.33\%}$$

However, this method is inaccurate as it does not account for the proportion of polyps that progress to cancer in the interim period. In the case study evaluation, the exact incidence and prevalence of polyps is calibrated to observed estimates of the age-specific prevalence of polyps and preclinical cancers.

Polyp growth and transition to cancer rates

Some models specify a constant time to transition of polyps to cancer, for example, all polyps

transform to cancers 6 years after their incidence. This approach contradicts evidence showing variation in transition rates and a distribution of such rates would seem to be more appropriate. The case study evaluation estimated constant transition probabilities from low- to high-risk polyps, and from high-risk polyps to cancer based on evidence from a limited number of studies. The point estimates for both transition probabilities were allowed to vary during the calibration process, which reflected the uncertainty in the directly estimated parameter values.

Without the use of tunnel states, the case study implementation of a Markov model is restricted to the assumption of an exponential distribution of transition times, although the directly fitted time-varying Weibull distribution was found to be similar to the fitted exponential distribution for extrapolating transition probabilities from high-risk polyp to cancer.

Progression through undiagnosed cancer states and symptomatic presentation

There is no direct evidence of the rate at which undetected cancers progress through alternative cancer states and the rate at which cancers within each of these states present symptomatically. The case study evaluation defined wide ranges for each transition probability and calibrated a range of potential parameter values to the output data used in the calibration process, which is a similar approach to that used by almost all previous decision analytic models.

A more direct estimation approach was investigated, based on Bayesian approaches to evidence synthesis, using data from screening trials and national incidence data. Only preliminary estimates were derived, although this method is being developed to incorporate the use of additional data sources, such as audit data, such that it may provide a valid alternative approach to the general calibration of these parameters.

Test characteristics
The case study evaluation identified a range of potential methods for the estimation of screening test characteristics:

- 'screening' before and independently from diagnostic evaluation
- 'screening' patients with known colorectal cancer
- diagnostic work-up of test-positive individuals and the follow-up of test-negative individuals to identify missed cases

- diagnostic work-up of test-positive individuals and test-negative individuals, and the application of a work-up bias correction method
- combine the observed (age-specific) screen-detection rate with the (age-specific) estimates of prevalence.

Each of the above methods has limitations, and the most appropriate method will vary according to the characteristics of the disease, such as the lead time that influences the effectiveness of the work-up and follow-up method, and characteristics of the screening study, such as the selection procedure for diagnostic evaluation.

Cancer mortality
The categorisation of cancer into a small number of stages, such as the A to D categorisation common in CRC, and the subsequent assumption that all individuals in a particular category have the same prognosis, may be inaccurate as screen-detected cases may be detected earlier or later within different stages. Whynes and colleagues presented a linear regression analysis investigating the impact of stage at diagnosis on age at death, and a Cox regression analysis determining the hazard ratio by participation in screening.[132] Similar analyses that combine stage at diagnosis and screening participation could investigate the potential for differential prognoses between screen-detected and symptomatic diagnosed patients.

The case study evaluation identified an audit study based in Wessex, which reported survival over the period 1991–5. These data are not ideal as treatment options for all stages of cancer have changed significantly since that period. The timescale of the case study evaluation precluded a more detailed representation of mortality post-diagnosis using a predictive treatment model rather than a historical observational dataset. The development of integrated screening and treatment models would provide a number of advantages over the use of observational data to estimate mortality rates following diagnosis with cancer:

- Survival is based on current treatment options.
- Costs and QALYs can be estimated more comprehensively.
- Treatment models can be updated to assess the impact of new treatment options on screening.

There are two broad approaches to the integration of a treatment model with a screening model for CRC:

- Separate treatment models could be appended to the screening model that describe patient pathways from diagnosis with CRC Stages A, B, C and D to death.
- A single treatment model could be attached to the screening model that describes the possible progression of patients from Stage A through each cancer stage to death.

Published treatment models cover most cancer stages, although we are not aware of a combined treatment model that adequately facilitates the cost-effectiveness analysis of all cancer stages. Models for the evaluation of interventions for early-stage cancer[133,134] tend to describe the later cancer stages in significantly less detail than do models developed for the analysis of later cancer stages.[135,136]

The process of developing an integrated treatment model and adapting and updating the input parameter values would be a considerable undertaking. However, the combination of screening and treatment models would provide a valuable resource that could be used to model the optimal portfolio of screening and treatment options, which could be updated as new screening and treatment options become available.

Screening uptake rates

The review identified that few studies implemented differential uptake rates for alternative screening programmes. Differential screening uptake rates by screening programme can have a significant impact on the estimated cost-effectiveness of alternative screening programmes. *Table 8* presents a hypothetical example of the effect of alternative uptake rates for two screening programmes. The second column presents the mean costs incurred, and QALYs gained, by individuals who comply fully

with the different screening options. The costs and QALYs in the subsequent columns comprise weighted sums of the costs and QALYs presented in the second column; for example the costs associated with screening programme 1 at a 50% compliance rate are as follows:

$$\text{screening programme 1 costs at 50\% compliance} = \begin{array}{l}(0.5 \times \text{no screening costs}) + \\ (0.5 \times \text{screening} \\ \text{programme 1 costs}) = \\ (0.5 \times £500) + (0.5 \times £2000) \\ = £1250\end{array}$$

The incremental cost per QALY of each screening programme compared to no screening remains constant regardless of the proportion of the eligible population screened. The incremental cost per QALY moving from programme 1 to programme 2 also remains constant for equal uptake rates, but decreases significantly when differential uptake rates are assumed.

Uptake rates may be related to costs, disease prevalence and progression. The effect of uptake rates on costs is potentially the simplest to handle as it requires the estimation of the variable cost per test undertaken and the fixed cost of a screening programme, with the latter being divided over the varying number of screening tests for different uptake rates.

The relationship between uptake and disease prevalence and disease progression is less certain and more complex to model. In the area of CRC, it has been hypothesised that individuals who do not attend screening may have a higher cancer incidence that those who do attend, and that disease progression may be faster in non-attenders. Scholefield and colleagues[137] report on mortality in three cohorts from the Nottingham FOBT CRC screening trial: attenders of at least one screening

TABLE 8 *Example of the effect of different uptake rates across and between alternative screening programmes*

| | | Compliance | | | | | | |
| | | 50% programme 1 50% programme 2 | | | 75% programme 1 75% programme 2 | | | 50% programme 1 75% programme 2 | |
	Cost/QALY (£)	Costs (£)	QALYs		Costs (£)	QALYs		Costs (£)	QALYs
No screening (0)	500/25	500	25.0		500	25.0		500	25.0
Programme 1 (1)	2,000/25.1	1,250	25.1		1,625	25.1		1,250	25.1
Programme 2 (2)	3,000/25.12	1,750	25.1		2,375	25.1		2,375	25.1
(0) – (1) ICER		15,000			15,000			15,000	
(0) – (2) ICER		20,833			20,833			20,833	
(1) – (2) ICER		50,000			50,000			28,125	

TABLE 9 *Mortality rates summary by screening cohort from the Nottingham FOBT trial*[137]

Screen-related cohort	All-cause mortality rate per 1000 person-years	Colorectal cancer mortality rate per 1000 years
Screen attenders	2.14	0.55
Screen non-attenders	3.1	0.81
Total intervention group	2.56	0.66
Non-invited to screening (controls)	2.57	0.74

round, non-attenders invited to screening and those not invited to screening (the controls). *Table 9* presents the summary results for all-cause and CRC-related deaths in the three cohorts, showing that non-attenders had higher mortality from CRC and from all causes than either the 'acceptors' or the controls. The comparison of CRC death rates between the non-attenders and the controls implies that either the incidence of CRC is higher or that the prognosis of diagnosed CRC is worse. Poorer prognosis may be due to non-attenders presenting later, more aggressive, faster growing tumours, or poorer access to appropriate health care. The comparison of all-cause mortality rates implies that non-attenders may have less to gain from screening as they are more likely to die earlier from other causes.

The observed differences in the all-cause mortality rate between screening attenders and non-attenders can be handled by applying alternative all-cause mortality rates to the two groups in a model. Weights applied to the aggregate all-cause mortality rate can be estimated such that the relative risk of death between the two groups is replicated. Based on the non-attender:attender relative risk estimated from *Table 9* of 1.45 (3.1/2.14), the relative risks applied to the aggregate all-cause mortality rate would be 1.18 and 0.82 for non-attenders and attenders, respectively. The handling of variable CRC-specific mortality rates by attendance could be handled in a similar manner, by estimating relative risks to be applied to stage-specific survival rates post-clinical diagnosis and post-screen-detected diagnosis.

If the relationship between disease prevalence and progression is restricted to ever-attenders and non-attenders, limited additional structural complexity would be required. If mortality rates are linked to specific screening attendance patterns, it may be necessary to update the relative risk of the surviving population after each screening round, which would probably become infeasible using a cohort Markov model. An

individual sampling approach could handle the additional complexity with relative ease by attaching attributes to individuals describing screening attendance patterns.

Costs

The process of estimating costs as inputs to a screening model is similar to other economic evaluations. The search for appropriate data sources includes trials, observational studies and registers that may have collected relevant resource use information, to which current resource use estimates can be applied. Most evaluations incorporated secondary cost estimates based on published values, if not resorting to the use of expert opinion to provide resource use estimates.

There are two cost issues with specific relevance to screening. First, an accurate screening model should differentiate between the fixed and variable costs of the screening programmes included in the evaluation. The estimation of these costs will likely require the development of a cost model, ideally in consultation with individuals running a relevant screening programme or screening trial. The cost model should use a bottom-up approach to estimate the quantity of different elements of resource that would be required to provide alternative screening programmes, which can then be categorised as either fixed or variable cost elements.

Second, costs of disease management post-diagnosis should be estimated using an integrated treatment model, rather than the estimation of survival based on observation datasets and the application of a constant cost across the survival period. The use of alternative health states describing the pathway of individuals diagnosed at different cancer stages would allow the more accurate application of cost estimates to each relevant health state.

Utility values

The estimation of utility values to populate cost–utility screening models is covered in detail in

Chapter 9. Issues around the estimation of utility values are similar to those around the estimation of utility in other areas of economic evaluation. Two issues are of specific relevance to screening. First, it is often stated that the process of screening itself has an impact on the utility of participants (and possibly non-participants), although this impact has only been explored using hypothetical utility decrements in sensitivity analyses.[138] This aspect of utility may be important, for example, if a utility increment is demonstrated for negative test results or if there is a significant utility decrement for false positives and a low specificity rate. It is generally not possible for screening model analysts to undertake the primary research required to inform such utility values, so additional external research may be required to inform screening models.

Second, the description of utility post-diagnosis would be more accurately implemented by the use of a treatment model that described the progression of patients through progressive health states, to which alternative utility values could be applied.

Calibration

The case study calibration approach followed previous screening models that fitted unobservable parameters to outputs estimated by the full screening model, that is, the full model was populated, with ranges specified for those parameters that required fitting. The full model was then analysed over a large number of iterations in which alternative combinations of parameter values were tested. Unlike other modelling studies, a base case set of parameter values was not specified, but rather a probabilistic sensitivity analysis was undertaken using a subjectively defined group of input parameter sets.

More complex model structures incorporate a greater number of unknown parameters, which may lead to a greater number of parameter sets that appear to fit published incidence and mortality estimates. One of the key benefits of retaining a simpler model structure is the minimisation of unknown parameters that require calibration, which may result in increased confidence that the chosen set of parameter values is valid.

It is likely that there will be correlations between output parameters used in the calibration process for most screening models, such as between incidence, stage at diagnosis and mortality. One implication of such correlations is that the degrees of freedom against which the χ^2 statistic is estimated should be reduced.[69]

Correlation between input parameters is important when a probabilistic sensitivity analysis is undertaken, and may be addressed by sampling sets of input parameters that have been calibrated rather than sampling values from independent probability distributions for separate input parameters.

Chapter 5

A review of models for the cost-effectiveness analysis of screening for diabetes and cardiovascular disease

Introduction

The scope of the screening modelling literature relating to diabetes and cardiovascular disease is very broad, including different types of primary prevention (screening that identifies risk factors for disease), secondary prevention (screening that identifies asymptomatic disease) and tertiary prevention (screening that identifies early, potentially asymptomatic complications of an existing clinical condition). The range of modelling types used to evaluate interventions is also varied with examples ranging from simple decision trees to complex Monte Carlo Markov simulation models. This chapter discusses selected examples of modelling studies to illustrate the range of potential approaches. Examples of the choices made in developing cost-effectiveness models in these disease areas are highlighted and some key issues that influence the usefulness of models to decision-makers are identified.

There has been an increasing interest in screening for diabetes and cardiovascular disease as intervention studies (largely RCTs) have identified effective interventions for both risk reduction and early intervention. Most RCTs of interventions that reduce cardiovascular risk focus on identifying and treating individual risk factors. Diabetes RCTs have involved treating individuals with clinically diagnosed diabetes or treating impaired glucose tolerance (IGT) to delay progression to diabetes.

The literature searches identified a large number of modelling studies that evaluated screening relating to diabetes and cardiovascular disease. These searches also revealed some limitations of using 'screening' (and related terms) as a key search term as much of the coronary heart disease (CHD) literature used the terms 'primary prevention' and 'secondary prevention', which may imply a screening intervention (identification of asymptomatic risk factors for disease) without using the term. Further papers were identified from the reference lists of modelling and economic evaluation papers. Identified papers were classified according to the clinical field (diabetes or cardiovascular disease), the disease stage at which screening was being considered and what was being identified by the screening test. Examples from the main areas in which screening models were identified are described in *Table 10*.

Diabetes-related screening models

Examples of modelling methods, including the structure, assumptions and parameters, are

TABLE 10 *Diabetes and cardiovascular disease screening categories*

| Disease stage | Clinical state identified by screening test | |
	Diabetes	Cardiovascular disease
Risk factors for disease present ('primary prevention')	Obesity IGT	Hypertension (blood pressure) Hypercholesterolaemia (cholesterol level) Overall cardiovascular risk
Early (asymptomatic) disease present ('secondary prevention')	Type 2 diabetes	Abnormal ECG (on exercise testing for CHD) Asymptomatic AAA Asymptomatic carotid stenosis
Clinical (symptomatic) disease present ('tertiary prevention')	Diabetic retinopathy Diabetic renal complications	Abnormal carotid Doppler after stroke or transient ischaemic attack Angiography in polycystic kidney disease

discussed for primary prevention of diabetes, type 2 diabetes screening and retinopathy and renal complication screening. As might be expected, there are generally better data available for the natural history of complications in individuals who already have an established clinical condition. This may explain why there has been much more model development around retinopathy screening in diabetic populations than around screening for type 2 diabetes.

Primary prevention of diabetes

Segal and colleagues modelled the cost per life-year saved (and cost per diabetes-year avoided) for primary prevention of diabetes, based on the identification of individuals at high risk of diabetes, due to IGT, previous gestational diabetes or obesity, in order to intervene and reduce risk.[139] Markov modelling was used to model transitions between normal glucose tolerance, IGT and type 2 diabetes, using probabilities derived from intervention studies. Differential uptake of screening tests was not modelled and costs of screening were not always included where screening for diabetes and IGT was regarded as part of routine clinical management. Compliance with interventions was assumed to vary, with those not complying having the same outcomes as the control cohort.

Type 2 diabetes screening

The CDC Diabetes Cost-Effectiveness Study Group model estimated the cost-effectiveness of screening for diabetes in a US population,[140] and has subsequently been adapted by others.[141] A Monte Carlo simulation model was developed to examine the cost-effectiveness of intensive clinical management, and was then adapted to estimate the cost per QALY gained of a one-off prevalent screening round. Progression through complication modules is modelled as a function of duration of diabetes, ethnicity and glycaemic control (measured by HbA_{1c}).

The model assumed that diabetes is present for about 10.5 years before being clinically diagnosed, based on extrapolating the progression of diabetic retinopathy for the proportion of cases already having background retinopathy at diagnosis. Direct evidence for the usual duration of diabetes before clinical diagnosis was not available and other indirect evidence, such as the prevalence of undiagnosed and diagnosed disease at different ages, was not used to validate this assumption. Transition probabilities were directly derived from intervention and cohort studies in which individuals with clinical diabetes were followed,

with the "most appropriate" model parameters being selected by "an expert panel convened for the study".

The sensitivity analysis showed that the cost-effectiveness of screening was highly dependent on the effectiveness of earlier treatment, which was assumed to be the same as for clinically diagnosed cases. The question of subsequent screening rounds was not addressed, which may be problematic for policy makers since the conclusion is that screening once may be cost-effective but gives no information to support decisions about when to re-screen those who screen negative or who have impaired glucose tolerance on screening. A similar model using population data from Taiwan suggested a 5-year screening interval was cost-effective from age 30 years, assuming that incidence of type 2 diabetes is over 1%.[141]

A recent review compared the reported cost-effectiveness of intervening at different stages in the disease process and concluded that there was better evidence to support primary prevention and intensive treatment after clinical diagnosis than to support screening and earlier diagnosis.[142] None of the models reviewed allowed interventions at the different stages of disease progression to be directly compared. This may be a significant limitation for policy makers who may need to consider not only the relative cost-effectiveness of different primary prevention interventions but also the relative allocation of resources to primary prevention, screening and clinical management and surveillance of clinically diagnosed cases. None of the published models would allow these types of trade-offs to be considered without further development.

Diabetic retinopathy screening

Most of the diabetes screening modelling literature concerns retinopathy screening. These models address the general question 'what are the costs and the benefits of early detection of diabetic retinopathy?' Although the details of models vary widely, all studies broadly agree that screening is cost-effective (if not cost saving) and that the higher the underlying risk (which is directly related to the duration of diabetes and the level of glycaemia) the more cost-effective screening is and the shorter the optimum screening interval.[143–153] Current UK policy is based on a model developed specifically for the UK National Screening Committee by James and Little.[154]

Javitt and colleagues developed a complex Monte Carlo simulation incorporating decision trees and

Markov processes referred to as PROPHET (PROspective Population Health Event Tabulation).[143,144] These models use data from cohort studies and intervention trials to estimate disease detection and progression rates, assuming uptake of screening was 100% in the screened group and that individuals with retinopathy would not be treated unless they were screened. QALYs were used for the first time in 1996.[149] Fendrick and colleagues applied a similar model to a Swedish population with type 1 diabetes and used Swedish costs.[147]

Dasbach and colleagues modelled three disease states (low risk, high risk and blind) in a time-varying state-varying Markov model, and extrapolated annual transition probabilities from 4-yearly cohort data.[145] Their estimates of uptake and compliance were taken from the cancer screening literature and so may be rather low (65% for screening, 79% for follow-up) for a population which already has a clinical diagnosis. The main outcome measure is sight-years saved and costs include both the healthcare costs and the welfare costs of blindness-related disability (hence the result that screening can be cost-saving).

Brailsford and colleagues[155] and Davies and colleagues[153] estimated the benefits of screening using an individual sampling Markov model that was based on the patient-orientated simulation technique (POST) technique, which was originally developed to model patient flows. The key differences in the Brailsford model were that there was some opportunistic screening in the absence of an organised screening programme and compliance varied with gender and ethnicity. This led to the conclusion that screening of all cases with non-insulin dependent diabetes mellitus (type 2 diabetes) may not be as worthwhile as targeted screening.

More recent modelling exercises focused on the choice of screening interval. A Wessex Development and Evaluation Committee report reviewed other models and used the POST model to examine the incremental cost-effectiveness of changing the screening interval from 2 years to 1 year.[150]

Although the model types used are similar (largely simulations using Markov processes to reflect transitions between retinopathy states) and the baseline parameters are largely derived from the same cohort [from the US-based Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR) cohort], the different model

assumptions lead to varied findings. The major differences in model assumptions include the following:

- *Complexity of disease process*: particularly the number of different states and whether parameters have fixed values or distributions, or variables are correlated with other variables. Some models only include proliferative retinopathy states, some also model maculopathy states. The model developed for the UK NSC was relatively simple,[154] whilst the Crijns' model[152] was much more complex, including variable transition probabilities and incidence rates, variable sensitivity/specificity for different degrees of retinopathy, variable effectiveness of intervention, allows regression of retinopathy.
- *Screening options*: incidence screening model with multiple screening cycles until death or sight-threatening retinopathy,[148] screening interval fixed[143] or dependent on clinical diagnosis/glycaemic control/presence of background retinopathy,[144] or prevalence screening.[153]
- *Uptake*: screening participation and treatment compliance was either estimated or assumed to be 100%;[144,151] one study assumed that loss to follow-up for screening may be directly related to frequency of screening, i.e. more frequent screening will reduce loss to follow-up.[146]
- *Comparison intervention*: screening strategies using a range of technologies were compared with no identification and treatment of retinopathy (usual assumption) or with opportunistic identification and treatment (a more realistic assumption).[155]
- *Discounting of costs and health benefits*: some discounted neither,[143] some discounted only costs[145,147] and some discounted both costs and benefits.[149]
- *Treatment benefit*: treatment benefit was assumed to start either when an individual lost sight without treatment[150] or immediately from time of treatment (which may be simpler to model but will overestimate the benefit of screening).[143]
- *Criteria to exit screening programme*: individuals may remain in the screening programme until background retinopathy develops or until treatable retinopathy develops and this will strongly influence the appropriate screening interval.[150]

In contrast to the wide variation in model assumptions, the sources of model parameters tended to be similar and based on relatively few primary studies, as described in *Table 11*.

TABLE 11 *Data sources for key parameters in models of screening for diabetic retinopathy*

Parameter	Data sources
Incidence	US cohort data
Progression	Cohort and trial data (Wisconsin Epidemiologic Study of Diabetic Retinopathy cohort)
Detection	Trial data (Diabetic Retinopathy Study Research Group)
Test characteristics	Comparative screening studies
Effectiveness of intervention	Trial data (Diabetic Retinopathy Study Research Group)
Uptake data	Extrapolated from other conditions or local data from real programmes[155]

Diabetic renal disease screening

Relatively few papers have considered the cost-effectiveness of screening for renal disease risk, despite the evidence from clinical trials that blood pressure reduction, particularly using an angiotensin-converting enzyme (ACE) inhibitor, can reduce the risk of progression. An early paper that describes itself as a cost–benefit analysis only compares the cost per case detected for different laboratory albuminuria tests (Micral-Test and radioimmunoassay).[156]

Borch-Johnsen used a Markov chain in combination with a second-order Monte Carlo technique to describe the impact of renal disease in type 1 diabetes, assuming no intervention with a strategy of annual screening and treating microalbuminuria with an antihypertensive agent (generally an ACE inhibitor).[157] Hazard functions for the natural history of renal disease are calculated from a Danish cohort followed until 1984 (before screening was possible), assuming an annual increase in albumin excretion of 20% and extrapolating back from the onset of proteinuria to the estimated onset on microalbuminuria. Uptake of screening and treatment (started 5 years after diagnosis) is assumed to be 100%. The results are presented as a threshold analysis: costs of screening are balanced by savings from reduced renal disease if the treatment effectiveness is at least 10%.

Different conclusions are reached by Kiberd and Jindal, who compared the costs and QALYs from microalbuminuria screening with hypertension and macroalbuminuria screening in type 1 diabetes.[158] The model identified key parameters that determine cost-effectiveness: effectiveness of early treatment, impact of treatment on quality of life, test characteristics, incidence of nephropathy and associated hypertension, rates of uptake and loss to follow-up. By choosing a different alternative to microalbuminuria screening (effectively screening for a different risk factor rather than no screening), they highlight how much the incremental cost-effectiveness of a screening strategy depends on the alternative with which it is compared.

Cardiovascular screening models

Cardiovascular risk factors (particularly blood pressure and lipids) have been the subject of numerous cohort studies and randomised trials and so there is scope for modelling the impact of identifying and treating a range of asymptomatic risk factors in populations with varying baseline risk. There are also data describing the effectiveness of screening in populations who are at high risk because of pre-existing disease. The more difficult screening scenarios to model are those involving earlier diagnosis of asymptomatic cardiovascular disease. Even in the absence of robust data about the natural history, the relatively high prevalence of these conditions has encouraged a number of such scenarios to be modelled. The following sections describe different categories of cardiovascular screening models.

Primary prevention of cardiovascular disease

A major issue for cardiovascular disease risk modelling is establishing the relationship between individual (or combinations of) risk factors and health outcomes.[159,160] For example, an evaluation of a risk reduction programme targeting deprived US women used a model that predicted the 10-year probability of CHD [outcomes including angina, myocardial infarction (MI) and CHD death] based on age, total and high-density lipoprotein cholesterol, systolic and diastolic blood pressure, smoking status and diabetes status.[159] Information on changes in risk factors associated with the intervention were entered into the model to estimate the associated changes in the 10-year probability of CHD. For a given 10-year probability of CHD, the conditional life expectancy was calculated, which informed the

incremental increase in life expectancy associated with a specific intervention. A comparison of different strategies for familial hypercholesterolaemia screening used a similar decision analysis and life table analysis to estimate the life-years gained as a result of screening.[160] Instead of estimating the impact of interventions based on treatment trial data, this model used mortality data from a cohort of patients with familial hypercholesterolaemia before and after the introduction of statins and therefore reduced the number of assumptions made about uptake of treatment and treatment effectiveness.

Marshall and Rouse used a simple spreadsheet model to explore two important issues: who should be screened for cardiovascular risk and what interventions should be used in screened individuals?[161] This model described a hypothetical general practice population reflecting an average risk factor distribution to which the Framingham risk equation was applied to calculate each patient's risk of a cardiovascular event. Relative risk reductions associated with different intervention strategies were applied to estimate the impact of intervention. Subsequent correspondence on the *BMJ* website showed that where questions were raised about model assumptions it was often relatively straightforward for the author to demonstrate (and sometimes for readers to recreate the model and demonstrate for themselves) the impact on the results in a way that is not possible when published papers are based on complex disease models that are not in the public domain.

Moskowitz and colleagues[162] produced a multistage screening programme model that aimed to reduce misclassification error due to variability in individual tests for hypertension, although some of the key assumptions may be questions. Such assumptions include that treatment is only of benefit to those with persistently raised blood pressure on repeat testing ('true positives') and that treatment is only effective above a specific blood pressure threshold, whereas the evidence for the effectiveness of antihypertensive agents is based on treatment after a limited number of one-off readings. Not only is there no evidence that there is a cut-off below which treatment has no benefit, but the relative risks and benefits of treatment depend on overall risk of cardiovascular disease, not only the degree of hypertension. Closer collaboration between modellers and epidemiologists might reduce the risk of models being developed that are statistically sophisticated but of little practical use to practitioners and policy makers.

Secondary prevention of cardiovascular disease

Earlier detection of cardiovascular disease has been modelled for CHD,[163,164] carotid stenosis,[165–167] intracranial aneurysms[168,169] and abdominal aortic aneurysm.[170–174] One of these models directly compares a secondary prevention intervention (early detection of CHD) with primary prevention (treating hypercholesterolaemia with statins).[163]

CHD screening
Sox developed a simple decision analytic model to investigate the increase in life expectancy and costs associated with exercise testing and subsequent bypass surgery for asymptomatic CHD.[164] Although key model parameters, in particular the prognosis of asymptomatic disease and the impact of intervention, were based on inadequate data, the model consistently makes assumptions that favour screening. This systematic biasing of the model to favour screening allows the authors to claim confidently that modelling shows that screening is cost-ineffective despite the favourable assumptions. Such modelling may be helpful to decision-makers even in the absence of directly relevant data sources to inform model development. Junod used a simple decision tree to compare exercise testing as a screening test for CHD, preventive treatment for all hypercholesterolaemia patients and no treatment.[163] As with Sox's model, a major limitation was the lack of direct evidence on prognosis or the impact of treatment on asymptomatic disease. A strength is that the decision tree is fully reproducible as the model parameters are listed, though it is not explicit which value was based on which reference.

Carotid stenosis screening
Three models of screening for asymptomatic carotid atherosclerosis, all published between 1996 and 1998, were reviewed. Derdeyn and Powers compared prevalence and incidence screening in a "high prevalence" and "low prevalence" cohort using a spreadsheet model.[166] Life-table data were used to calculate mortality rates. Sensitivity analyses identified the effectiveness of intervention and the discount rate as the most sensitive variables and suggested that a "one-off" screening programme in a "high prevalence" population might be cost-effective, whilst annual screening would be detrimental due to increased intervention-related harm. Lee and colleagues used a decision tree model with Markov subtrees to address a similar question and also found, despite using different modelling techniques and

parameters, that cost-effectiveness was highly sensitive to the effectiveness of the intervention (more precisely the duration of survival benefit).[165] Yin and Carpenter similarly used a decision tree with Markov subtrees and the same RCT evidence for treatment effectiveness (the ACAS trial).[167] The paper also cited the previous models and compared their assumptions and conclusions, identifying differences in the assumed duration of benefit from treatment (5 years versus lifelong) as the primary reason for their different findings. The value of this paper is also enhanced by the presentation of the model structure including the Markov subtrees, values for all the model parameters used with the source references and the range used for sensitivity analyses.

Intercranial aneurysm screening

Two Japanese studies evaluated cost-effectiveness using decision trees. Baba and colleagues used a relatively simple decision tree to compare screening versus no screening.[169] Minimal sensitivity analyses were presented (only varying screening test characteristics) and the outcomes were life-years saved. Baba's model was not cited by Yoshimoto and Wakai, who used a decision tree analysis with Markov subtrees for each option (aneurysm detected by screening, screening and no aneurysm, no screening).[168] Assumptions and sources of parameters were reported in some detail, as were the results of sensitivity analyses. Outcomes were reported as cost per QALY and a key finding of the sensitivity analyses was that benefit only outweighed harm with the higher estimates for the rate of rupture in the absence of intervention. The implication was that although current evidence may not support the introduction of a population screening programme, it is imperative to collect more information on the rate of rupture in large cohorts to ensure an appropriate policy change if surgical techniques improve or rupture rates turn out to be higher in specific populations.

Abdominal aortic aneurysm screening

A number of modelling studies using different methods (from spreadsheet to Monte Carlo simulation, mainly published between 1993 to 2001) with very different conclusions were followed by an economic analysis undertaken alongside a clinical trial of screening that was extrapolated to estimate cost effectiveness at 10 years. Further modelling work based on the trial is ongoing. Given the potential for comparing the relationship between modelling approaches and model findings and for potentially using the RCT results to evaluate retrospectively the validity

of earlier models, we have used AAA screening as a separate case study (reported in Chapter 6).

Tertiary prevention of cardiovascular complications

The literature searches identified a number of studies that evaluated the use of screening tests in the management of specific conditions. These models were considered separately because some of the modelling issues, particularly the appropriate choice of comparator, will be related to the population being screened. Some authorities[175] would also debate the semantics of whether these are 'screening' models or are more appropriately described as 'clinical surveillance' or 'diagnostic testing' models as they are all in populations with a defined clinical condition rather than healthy, asymptomatic populations, even though they may not have symptoms related directly to the condition for which they are being screened.

Screening for deep vein thrombosis after total hip replacement

Sarasin and Bounameaux[176] used a decision tree with Markov processes to compare screening after 1–2 weeks anticoagulation with strategies of only treating for 1–2 weeks, treating for 6 weeks or treating for 3 months. Use of ultrasound or venography for screening was also compared. Parameters were all derived from published sources (trials and observational studies). Risk of major outcomes and costs were calculated for a hypothetical cohort of 10,000 patients and reported as a cost per pulmonary embolism averted. The reported outcomes did not allow a quantifiable trade-off to be made between an increase in major bleeding complications versus a decrease in the number of pulmonary embolisms in the cohort.

Screening for stroke risk in management of symptomatic carotid stenosis

Derdeyn and colleagues[177] used a decision tree with Markov processes to evaluate screening patients who already had symptoms in order to identify those at higher risk of a further stroke. The model compared screening [using positron emission tomography (PET) scanning] followed by surgical intervention in those at higher risk with usual medical management. The model combined data from a study of the use of PET scanning to predict stroke and data from a trial of the surgical intervention (extracranial to intracranial arterial bypass) to describe the effects of surgical intervention versus medical management. It is an example of a model being used to support the

need for a trial to evaluate an intervention. The model suggested screening could be cost-effective if surgery is as effective in the screened population as in previous less highly selected populations.

Screening for coronary heart disease before vascular surgery

Glance[178] used a decision tree with Markov subtrees to compare screening and surgery versus medical management in a population at high risk of adverse outcomes. Parameters came from previous models and clinical databases (patient cohorts). Probabilistic sensitivity analysis informed CIs around the model outputs.

Screening for intracerebral aneurysms in polycystic kidney disease

Butler and colleagues[179] used a decision tree with Markov processes to assess the cost-effectiveness of screening compared with no screening. Parameter values were derived from cohort studies. Results were reported in terms of gain in total survival and gain in survival without neurological disability. Costs were reported as the total costs to society of screening or non-screening strategies for a 20-year-old. The findings were reported to be sensitive to the prevalence of aneurysms, probability of rupture and probability that rupture is fatal or disabling. The paper described in some detail the potential causes of bias in model parameters.

Screening for carotid atheroma after stroke or transient ischaemic attack

Jespersen and colleagues[180] calculated the costs of screening and subsequent surgery, but gave no estimates of the related benefits. It should be possible to estimate the benefit of screening and surgery given that this has been done for asymptomatic populations.[165–167] The lack of modelling studies is presumably related to the existence of data from clinical trials of screening from which cost-effectiveness can be directly estimated without any need for modelling, although modelling might still be useful if comparing alternative screening options.

Discussion of issues common to diabetes and cardiovascular screening

When the purpose of screening is detection of asymptomatic disease, the choice is usually between screening (in a specified population) and not screening. Screening for AAA is an example where the appropriate comparison may be

between a screening programme which offers screening to an eligible population and no screening. Asymptomatic aneurysms will still be diagnosed when detected by an abdominal examination or ultrasound scan as an incidental finding in a 'no screening programme' scenario, but this will be a relatively small subgroup of those included in a screening programme. However, when screening is identifying complications of an existing condition [diabetic retinopathy screening or postoperative deep venous thrombosis (DVT) screening] the appropriate comparison may not be screening versus no screening. It may be a systematic screening programme versus clinical surveillance as part of routine care which may have lower sensitivity than a screening programme but will detect a significant proportion of cases (e.g. retinopathy screening). Alternatively, the appropriate comparison may be screening, diagnosis and selective treatment of a complication versus universal treatment (e.g. microalbuminuria screening versus universal treatment with ACE inhibitors, or DVT screening versus universal anticoagulation therapy).

If models assume no screening in the absence of an organised programme, it will overestimate the impact of a screening programme in a scenario where there is already opportunistic or 'haphazard' screening (the most likely policy scenario for diabetes screening). Similarly, a model that assumes that uptake of screening is 100% and compliance with treatment is 100% will lead to overestimates of the impact of a screening programme.

Screening test results are often dichotomised as 'positive' or 'negative'. This is virtually always an oversimplification. The risk of disease after an abnormal screening test depends on the pre-test risk and the exact nature of the abnormal screening test. When screening is identifying risk factors for cardiovascular disease or diabetes, which are often themselves continuous variables (e.g. blood pressure, blood glucose, blood cholesterol, body mass index), it is more likely that models will need to reflect the actual distribution of risk associated with different risk factors. More recent risk tables and 'risk engines' are designed to quantify individual risk based on risk factors and can be used to model the impact of screening for, and treating, a range of risk factor values. Models in which treatment pathways in diabetes or cardiovascular disease are determined by one dichotomised risk factor (e.g. blood glucose level defined as diabetes or not, blood pressure level defined as hypertensive or not hypertensive[162])

will not reflect the reality of clinical decision-making.

The models reviewed in this chapter often assumed that screening only occurs once, either for a same-age cohort[166] or a cohort with the age distribution of a typical general practice population.[161] This is appropriate for conditions that have an incidence of close to zero in a screened cohort and for which the screening test has high sensitivity, suggesting that there will be few true positives detected by a subsequent screening round. It may also be appropriate if modelling demonstrated that a prevalent screening round could not be cost-effective, as it is then unlikely that repeat screening could ever be cost-effective. If neither of these two conditions hold, then it is important to consider whether modelling of further screening rounds is appropriate, as has been done for retinopathy screening but not for diabetes screening.

The reviewed models of secondary prevention (screening) for diabetes and cardiovascular disease generally contained less detailed representations of disease natural history than the models of screening for cancer that were reviewed in Chapter 2, although the post-diagnosis pathways were by and large more complex. The natural history models did not tend to describe progression through different preclinical states. Whereas most cancer screening models calibrated unobservable natural history parameters to datasets to derive the 'best fit' model parameters, the authors of cardiovascular and diabetes screening models were generally more willing to use expert opinion to estimate unobservable parameters. There was also a lack of information about the external validation of models, even in fields such as cardiovascular risk screening or diabetic retinopathy screening, where observational data from screening programmes could be compared with the results from modelling exercises. Most of the literature reported the results of a specific model with only passing mention, at most, of how the assumptions or results differ from other models in the same field. There was also often a lack of detailed information about the structure, assumptions and parameters of the applied models which made it difficult to compare across models or understand the underlying reasons for very different conclusions.

The following sections summarise the key issues identified in this chapter that may affect the usefulness of models in informing decision-making.

Identifying the key questions

Identifying the key clinical or policy question(s) that a model is intended to inform should be the first step in model development. The most useful model is arguably the one most directly relevant to decision-making rather than the one that makes the most elegant use of available data. Consultation at an early stage with clinicians, policy makers and public health professionals should improve the relevance of models without loss of methodological rigour and encourage dialogue between model developers and model users which could help ensure models are used appropriately in informing decision-making. Screening models may be used to inform both individual clinical decisions ('Should this individual be screened for this condition?') and public health policy decisions ('Should we introduce a screening programme for this condition?'). In both situations, an appropriate cost-effectiveness model can provide useful information on the relative costs and benefits of different screening protocols and identify population groups or individuals for whom the benefits of screening are likely to outweigh the adverse consequences and identify those for whom screening may be a (relatively) cost-effective intervention.

Models for assisting individual decisions about screening can also incorporate parameters that depend on individual values or choices. This may include the values associated with information (e.g. whether an individual values information about future health risks above increased health insurance or life insurance costs) and the choices about interventions an individual could make on receipt of a screening test result or diagnosis.

Care needs to be taken in extrapolating from models based on clinical decision-making to assist public health policy making. Policy makers considering a new screening programme may have to consider the impact of varying uptake rates, additional costs of programme management and quality assurance, the need to train additional staff for the screening programme and also to staff the additional diagnostic and treatment services that will be required if screening is introduced. For example, the costs of a screening programme involving a surgical intervention, such as AAA screening, should include the costs of training the additional medical, surgical and nursing staff required and the additional surgical procedures and surgical and intensive care facilities.

Public health policy decisions may require models with a broader scope than a comparison of

screening versus not screening. For an individual the clinical question may be relatively narrow: is it worthwhile doing this screening test for this individual at this time? The associated policy question may be broader: is it more cost-effective to invest in primary prevention, in screening and earlier treatment or in more effective treatment of cases already diagnosed?

Appropriate presentation and comparison of models

Model structures need to be presented in sufficient detail for their limitations to be fully understood. Sources need to be explicitly stated and choices justified for model parameters, even where there is involvement of an 'expert panel'. Model limitations need to be discussed, particularly the key areas of uncertainty and how they will affect model results.

Current models should be compared with previous models, where published. They should discuss how and why their model structure, parameters and results differ from previous published models. This appears to be a relatively neglected area with scope for more robust comparison and critique of the multiple models that have often been developed in a single clinical area. Where these comparisons have been done it is extremely useful in assisting policy makers in identifying the relevance of specific models to existing screening scenarios or local populations. If differences in results are due to difference in the values chosen for key parameters, this can be demonstrated by giving the results when the same parameter values are used.

Chapter 6

The users' perspective: a case study assessment of screening for abdominal aortic aneurysms

Introduction

The objective of this chapter is to investigate further the issues raised in Chapter 5 about screening models from the users' perspective. Seven papers describing six alternative approaches to modelling the cost-effectiveness of screening for asymptomatic AAAs were identified and reviewed.[170–174,181,182] These models addressed a range of policy questions, incorporated alternative model structures and sets of assumptions and presented differing levels of detail about the respective models. This set of models provides a good basis for the review and critique of screening modelling methods from the perspective of the non-analyst user because they illustrate the range of issues that affect the usefulness of models for the user. The relevance of the comparison is furthered by the fact that the model results ranged from the prediction that screening for AAA is extremely cost-effective to the finding that screening for AAA is more costly and produces fewer life years than a no screening option (i.e. screening is dominated).

All of the reviewed models were developed prior to the publication of results from an RCT of screening for AAA, which included a corresponding economic analysis. The trial-based cost-effectiveness analysis provides a solid source for validation, in addition to informing an assessment of how the data from the trial could be used in a model-based evaluation.

The published models of screening for asymptomatic AAAs were reviewed in detail and alternative approaches to the following issues were documented:

- the relevance of the policy question(s) addressed
- the model structure and assumptions
- the choice of data sources and the presentation of sufficient information for the results to be interpreted in the context of alternative decision areas
- validation, including cross-validation with other models of screening for AAA.

Relevance of the policy question(s) addressed

This section discusses the relevance of the policy questions addressed by the reviewed studies, relating to the following factors:

- the screening strategy, i.e. the eligible population and the frequency of screening
- the threshold AAA size at which elective surgery is offered
- the threshold AAA size at which persons are kept under surveillance
- the specification of surveillance strategies
- the procedures in place for detecting AAAs before rupture in the absence of a screening programme.

Mason[171] evaluated the incremental cost per life-year gained of a policy option of offering one-off ultrasound screening to all men aged 70 years compared with no screening. A threshold for surgery of 5 cm was assumed. Patients with aneurysms between 3.5 and 5 cm were re-examined at 6-monthly intervals, which was assumed to prevent any ruptures in patients who were not contraindicated to surgery. The incidental detection of AAAs in the absence of screening was not modelled.

Frame and colleagues[170] presented the incremental costs per life-year gained moving through the following intervention options:

- no medical care
- emergency surgery only
- elective surgery for AAAs identified pre-rupture in the absence of screening
- one-time screening with ultrasound/physical examination for men aged 60–80 years
- screening men aged 60–80 years with repeat ultrasound/physical examination 5 years after the first screen.

All men in whom an aneurysm larger than 4 cm is detected would be offered surgery. Frame and colleagues recognised that the decision to undertake elective surgery was commonly based on a 5-cm threshold, but prevalence data were not available

for a 5-cm threshold. The size of AAA at which a patient enters surveillance was not specified, although a surveillance procedure was described that implied a zero probability of rupture whilst on surveillance. A proportion of AAAs was assumed to be detected in the absence of screening, although the process by which they were detected was not described.

Law and colleagues[181] did not specify a research question, although their analysis suggested that they were evaluating the mean cost per life-year gained (compared with no screening) for alternative surgery thresholds for AAAs for men aged between 60 and 80 years. From a policy perspective, the main limitation of this study is that it did not evaluate a specific screening programme; rather, it assumed that the necessary screening would be undertaken to identify all AAAs by the time they reached alternative thresholds for surgery. The incidental detection of AAAs was implicitly modelled through the data analysis methods (see the section 'Model structure and assumptions', p. 59, for a more detailed description).

St Leger and colleagues[174] reported that they examined screening elderly males for aneurysms of 5 or 6 cm to estimate the cost per additional QALY gained. Further reading of the paper reveals that a surgery threshold of 6 cm was assumed, and that patients with AAAs between 3 and 6 cm were followed annually by repeat ultrasound. Aneurysms could rupture under surveillance before they grew to the 6-cm surgery threshold, although not after they reached the threshold. No incidental detection of AAAs in the absence of screening was modelled.

Pentikainen and colleagues[172] evaluated the incremental cost-effectiveness of one-off targeted ultrasound AAA screening, for first-degree male relatives aged between 50 and 85 years of patients with AAA. A secondary analysis of screening both first-degree male and female relatives was also undertaken. A threshold of ⩾5 cm was used for the offer of elective surgery. AAAs sized between 3 and 5 cm were assumed to be re-examined every 6 months, and those between 2 and 3 cm annually. It was implied that AAAs could rupture at any point during surveillance, that is, before and after they reached the surgery threshold. The detection of AAAs pre-rupture in the absence of screening was modelled, although the process by which they were detected was not described.

The same group published a cost-effectiveness analysis of targeted screening for AAA in 2001,[173]

although this study evaluated screening first-degree male relatives only.

Lee and colleagues[182] evaluated one-time ultrasound or 'quick-screen' screening for 70-year-old male patients with a high prevalence of AAA. The 'quick-screen' appears to be a screen performed by ultrasound, but undertaken in a maximum time of 5 minutes. A 'quick-screen' was defined as 'limited' if adequate measurements could not be achieved within 5 minutes. High-risk individuals were identified using the following criteria: current or former smoker; hypertension; hyperlipidaemia, coronary artery disease; history of lower extremity bypass operation; claudication; ischaemic rest pain; and carotid artery disease.

A surgery threshold of ⩾5 cm was assumed, with AAAs sized between 4 and 4.9 cm being re-examined every 6 months, and AAAs between 3 and 3.9 cm annually. AAAs could rupture during surveillance, although it was implied that ruptures did not occur after an AAA reached the surgery threshold. A proportion of AAAs would be detected in the absence of screening, although the process by which they were detected was not described.

Comment on research question and policy variables

The overriding comment regarding the specified (or implied) policy questions in the reviewed modelling studies is that the individual studies did not utilise one of the principal benefits of a modelling approach: the evaluation of a range of alternative screening programmes. The evaluation of alternative policy options appears to be particularly relevant in an area such as screening for AAA, where there are numerous policy options for a range of parameters that are likely to affect the cost-effectiveness of screening.

Three policy parameters, in particular, were varied across the reported studies: the definition of the eligible screening population; the comparison of one-time screening with repeat screening; and the threshold for surgery.

The studies either defined the eligible population as men at a specified age (e.g. 70 years old), or between certain ages (e.g. 60–80 years old). The specification of an age range for the eligible population is unlikely to be policy relevant at the point of implementation of a screening programme, and certainly not at the point at which screening reaches equilibrium, i.e. when all persons within the age range have been screened

and the screening programme is screening people as they reach the lower bound of the age range.

The appropriate research question is a two-part question that first determines the optimal age at which the prevalence screen should be offered when the programme is operating in equilibrium. Second, the maximum age to which it is cost-effective to offer a prevalence screen at the point of implementation of a screening programme should be determined. Each part of the question is subject to complicating factors, such as the need to account for alternative repeat screening strategies when estimating the optimal age for the prevalence screen, or the need to account for resource limitations in the real world when estimating the scope of the initial prevalence screen.

Two studies defined high-risk subgroups at whom screening could be targeted, for example, Pentikainen and colleagues[172] evaluated screening first-degree male (and female) relatives of patients with identified AAAs, whereas Lee and colleagues[165] specified a series of high-risk characteristics. The broadest approach would be to specify all relevant combinations of risk characteristics by which an eligible population could be specified, for example, patients with hypertension, patients with hyperlipidaemia, patients with hypertension and hyperlipidaemia. Population-based screening should be included as policy makers may need to provide an economic justification for excluding this most equitable policy option.

The study reported by Frame and colleagues[170] is the only one that explicitly models the cost-effectiveness of repeat screening compared with a one-off screen. The exclusion of this option is justified if one-off (or prevalence) screening is not found to be cost-effective, as incidence screening can only be less cost-effective. However, most studies found one-off screening to be cost-effective, but still did not evaluate repeat screening. Few data describing incidence rates may have been available at the time of analysis of most studies, but further discussion justifying the exclusion of repeat screening as a policy option would better inform the user. Preferably, the model would evaluate repeat screening at alternative intervals. At least one recent clinical paper has estimated the yield of repeat screening,[183] and further options for incorporating such data into a model-based evaluation of repeat screening are presented below.

There does not appear to be a consensus regarding the appropriate AAA size threshold above which elective surgery is offered, which varies from 4 to 6 cm. It would appear reasonable to test the cost-effectiveness of AAA screening given alternative treatment thresholds. There was also some variation in the size at which an AAA qualified for further surveillance, including 2, 3 and <4 cm. These options, too, could be investigated.

The above comments illustrate the range of potential policy options that could have been evaluated. A good natural history model that differentiates between the incidence and progression of aneurysms at different ages (and preferably between different risk groups) facilitates the evaluation of all such options.

Model structure and assumptions

This section describes the structures of the AAA screening models and the assumptions made about the natural history of AAAs and the screening process.

Figure 10 shows the decision tree used by Mason[171] to represent the pathway of men who are screened and are candidates for urgent elective surgery (aneurysms larger than 5 cm detected), require follow-up or have no aneurysm detected. Men who were not contraindicated either underwent surgery or were followed up, with a proportion of the latter developing large aneurysms and undergoing surgery. The analysis compared the survival profile over the subsequent 10 years for men in whom an aneurysm was detected through screening (excluding those who were contraindicated for surgery and those in whom a small aneurysm did not become large, and hence did not require surgery). Estimates of the proportion of small aneurysms that became large and an annual rupture rate for large AAAs were used to estimate the number of ruptures in the no screening scenario. Differential adjustments to the all-cause survival rate for screened and non-screened post-surgery survivors were applied to the respective cases to estimate the life-years gained over the 10-year period.

The simple model structure described the screening and disease process in an appealing and intuitive manner, and is easily replicable. The decision tree is potentially appropriate because the study only considers a one-time prevalence screen. However, the chosen model structure required a
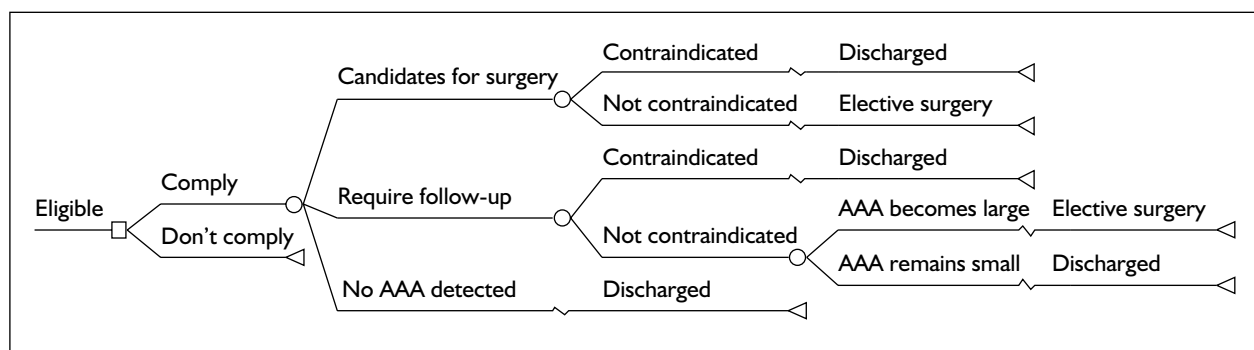
FIGURE 10 *Decision tree structure used by Mason[171]*

TABLE 12 *The impact and likely bias of model assumptions made by Mason[171]*

Assumption	Impact	Interpretation
Equal risk in attenders and non-attenders	Risk in attenders may be inaccurate if based on aggregate rates	May favour screening if non-attenders have higher risk
A constant transition probability between two AAA sizes	The age-specific distribution of AAA size in the smaller size state is not represented	Likely to favour screening, as transitions, and hence ruptures, will occur earlier in the model
There is a constant rupture rate for all large AAAs	The size distribution of large AAAs will vary by age	Likely to favour screening, as ruptures will occur earlier in the model
10-year time horizon	Follows men to the age of 80 years, so ruptures and survival post-80 years are not included	Probably favours no screening, as fewer screened men die as a result of AAA

number of simplifying assumptions that might compromise the validity of the evaluation. *Table 12* describes assumptions made in the model that may affect a user's interpretation of the results. The most problematic assumptions include the 10-year time horizon, which limits the benefits of screening, whereas the assumptions of no ruptures during surveillance and no opportunistic detection both favour screening. Unlike some modelling studies in which assumptions all favour one intervention, the direction of the assumptions in this model varied so it is not possible to interpret the impact of the assumptions qualitatively (i.e. the overall direction of the bias), let alone quantitatively (the magnitude of the bias).

No diagram of the model used by Frame and colleagues[170] was presented, but it was described as a spreadsheet model containing seven states: (1) no AAA; (2) undiscovered AAA smaller than 4 cm; (3) undiscovered AAA larger than 4 cm; (4) discovered AAA smaller than 4 cm; (5) AAA larger than 4 cm, which was discovered when less than 4 cm and not re-examined since discovery; (6) discovered AAA larger than 4 cm; (7) dead. The model moved forward in annual cycles over a

20-year time horizon, although the sequence of events within each cycle was carefully defined: screening, annual re-screening of small aneurysms, incidence of interval aneurysms, and ruptures occurred at subsequent days at the start of each year, whereas other-cause deaths, new aneurysms and the progression of aneurysms from small to large occurred at the end of each year. This set-up meant that all men under surveillance (state 5) whose AAA became large would be re-examined (and offered elective surgery) before the aneurysm ruptured. It also favoured screening as all ruptures occurred at the start of the year, thus having a shorter survival period than if assumed to rupture at the midpoint of the year.

The model allows for the diagnosis of AAAs by case finding in the absence of formal screening. *Table 13* describes the assumptions incorporated in this model, which include the specification of only two AAA size categories and a constant growth rate of initially small to large aneurysms. Persons surviving elective surgery or a rupture were stated to be cured, presumably entering a post-AAA state, from which there is an annual incidence of developing a new AAA. The specification of an

TABLE 13 *The impact and likely bias of model assumptions made by Frame and colleagues*[170]

Assumption	Impact	Interpretation
100% screening uptake	Not able to define differential risks for attenders and non-attenders	May favour screening if non-attenders have higher risk
A constant transition probability between two AAA sizes	The age-specific distribution of AAA size in the smaller size state is not represented	Likely to favour screening, as ruptures will occur earlier
Non-age specific prevalence of AAAs	Wide age range for screening defined	Likely to favour screening if prevalence rates are increased for younger persons (more life-years to gain)
No small AAAs rupture during surveillance	No data required to estimate such ruptures	Favours screening
20-year time horizon	Follows 60-year-olds to the age of 80 years, so ruptures and survival post-80 years are not included	Probably favours no screening, as fewer screened men die as a result of AAA

TABLE 14 *The impact and likely bias of model assumptions made by Law and colleagues*[181]

Assumption	Impact	Interpretation
100% screening compliance No ruptures during surveillance No contraindications to surgery	Detection rate based on size of AAA at ruptures	Favours screening, as screening prevents all ruptures occurring in AAAs above threshold size

eligible population between 60 and 80 years introduced some uncertainty as the age distribution of the screened population is not defined. The population of screened 60-year-olds is likely to have a very different prevalence of AAAs to the population of 80-year-olds and it is not clear how the clinical parameters accounted for the stated age range.

Law and colleagues[181] presented an equation-based model that estimated the number of lives saved by screening, for alternative surgery thresholds, as

$$\text{lives saved} = (1 - \text{elective operative mortality rate}) \times (\text{detection rate} \times \text{number of AAA deaths}) - (\text{elective operative mortality rate} \times \text{ratio of ruptured to unruptured AAAs at death} \times \text{detection rate} \times \text{number of AAA deaths})$$

The detection rate for a specified surgery threshold AAA size was estimated from data describing the distribution of the size of AAAs at rupture, that is, the detection rate is the number of AAAs that ruptured above the threshold size divided by the total number of ruptures. The ratio

of ruptured to unruptured AAAs at death was based on the proportion of AAAs above a threshold size that ruptured prior to death. The combination of the ratio of ruptured to unruptured AAAs at death, the detection rate and the number of AAA deaths informed the number of false-positive surgical interventions, that is, surgical episodes for persons with AAAs who would have died from other causes prior to rupture. An incidental detection rate for AAAs was implicitly incorporated as the analysis was based around the number of AAA deaths.

Table 14 lists the assumptions implied from Law and colleagues' study. The principal limiting assumptions appear to be that there were no contraindications to elective surgery, there was 100% compliance with screening and there were no ruptures during surveillance, which were all implied by the method used to calculate the detection rate.

St Leger and colleagues[174] used a Markov model with a time horizon of 5 years. The model structure was not explicitly presented, although the progression of men with an aneurysm in year 1 through a maximum of eight size-defined aneurysm states is described. Progression is based

on a size-specific annual AAA growth rate and annual risk of rupture. Men reaching the ≥6 cm category received elective surgery if screened, whereas non-screened men remained at risk of rupture. An 'other-cause' mortality rate was not incorporated, although it was stated that background mortality was indirectly introduced through the estimation of QALYs.

It is not clear how the results of the rupture model were used, but a best guess is that the number of electively repaired AAAs estimated by the model were assigned a normal age-specific life expectancy (assumed to be 12 years), and the sum of the life expectancies was then compared with the sum of the life expectancies for the same patients in the absence of elective surgery (assumed to be 2 years). It is not clear that the non-screened rupture model informed the estimation of QALYs at all. Perhaps the non-screened rupture model only informed cost estimates for the non-screened group.

Table 15 describes the main assumptions underlying the model. A particularly questionable assumption was the choice of a 5-year time horizon, which left over 92% of the original population of people with AAAs in the no screening arm still alive (which favours no

screening). The assumption that AAAs cannot rupture while under surveillance favours screening. The qualitative impact of these assumptions is mixed, which makes the interpretation of the results more difficult.

Pentikainen and colleagues[172] used patient-level DES to describe the pathway of persons through a screening model (*Figure 11*), and through current practice (*Figure 12*). This model used the most detailed breakdown of AAA categories. AAAs detected in the current practice model were assumed to follow the size-specific pathways described in the screening model. The assumptions behind the model are described in *Table 16*.

The screening model invited persons from a defined probability distribution of ages between 50 and 84 years. It appears that the assumption of an equal prevalence of AAAs across this age range was taken due to the small sample size in the study that informed prevalence estimates (a local empirical study). This assumption would have a greater impact if the model is used to inform continued screening at age 50 years once the screening programme reaches equilibrium (i.e. when screening is being offered to individuals as they reach 50 years). The assumption that no AAAs over 5 cm ruptured during surveillance

TABLE 15 *The impact and likely bias of model assumptions made by St Leger and colleagues[174]*

Assumption	Impact	Interpretation
5-year time horizon for rupture model	Incidence of ruptures is underestimated	Favours no screening, as screening would prevent more AAA deaths
Other-cause mortality not incorporated in the 'rupture' model	Men are not at risk of dying from other causes over the 5-year time horizon	Favours screening, as will overestimate number of ruptures
No 6-cm AAAs rupture during surveillance	No data required to estimate such ruptures	Favours screening
All AAAs grow by a size-specific amount every year	There is a mismatch of growth and AAA size categories	Unknown bias

TABLE 16 *The impact and likely bias of model assumptions made by Pentikainen and colleagues[172]*

| Assumption	Impact	Interpretation
Equal prevalence of AAAs across ages 50–84 years	Age-specific prevalence is not represented	Likely to favour screening, as more AAAs will be prevented at younger ages
AAA growth modelled as a continuous function	All AAAs assigned a specific size at the point of screening (in mm)	Intuitive representation of the disease process
AAAs over 5-cm threshold cannot rupture during surveillance	All 5-cm AAAs under surveillance are offered surgery	Favour screening
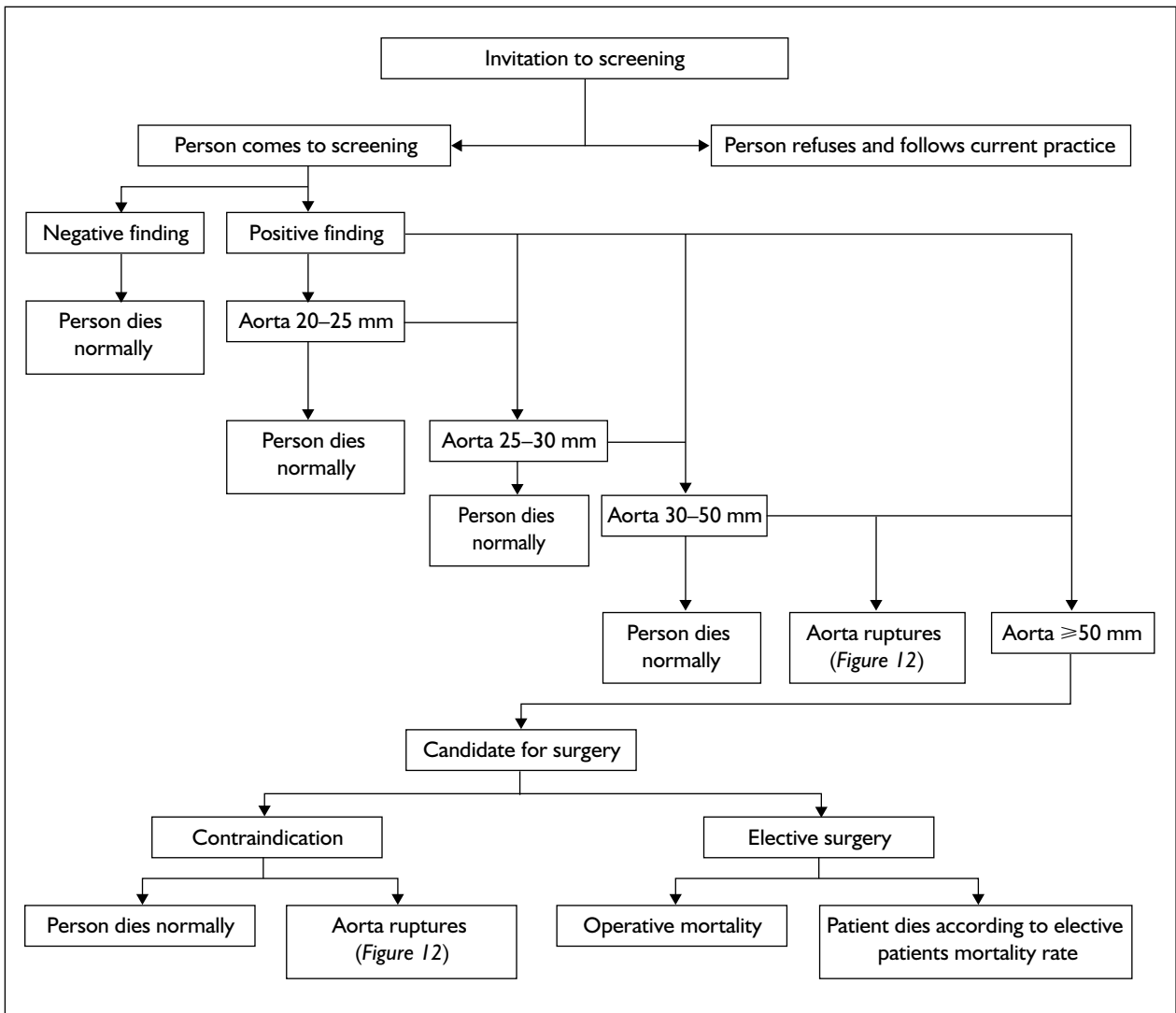
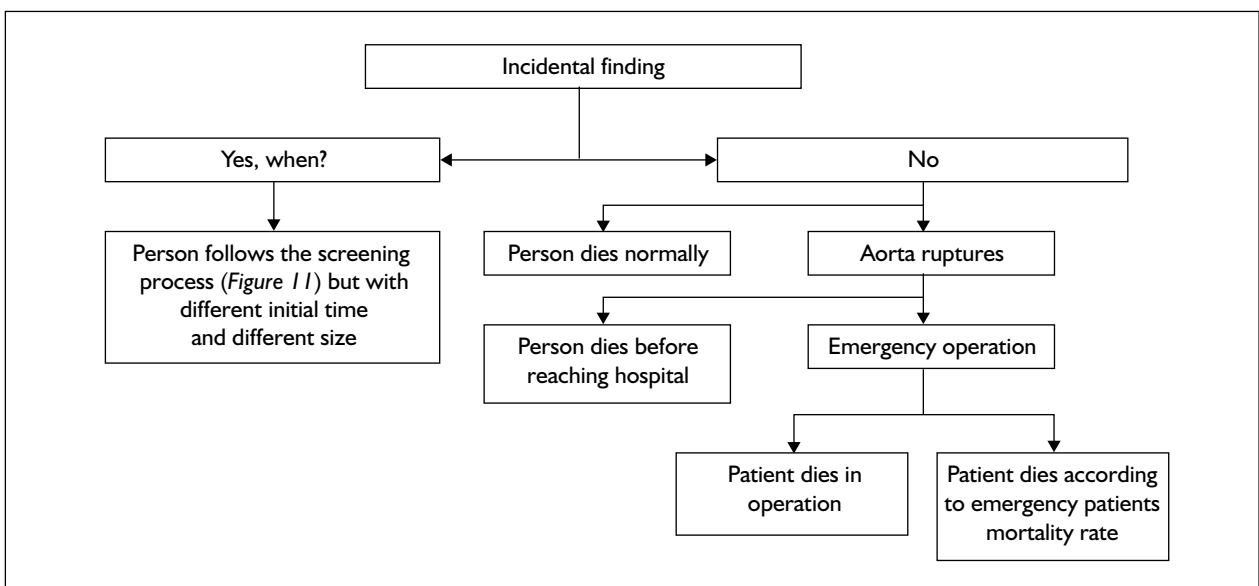**FIGURE 11** *Model structure for AAA screening presented by Pentikainen and colleagues[172]*

FIGURE 12 *Model structure for current practice in the absence of AAA screening presented by Pentikainen and colleagues[172]*
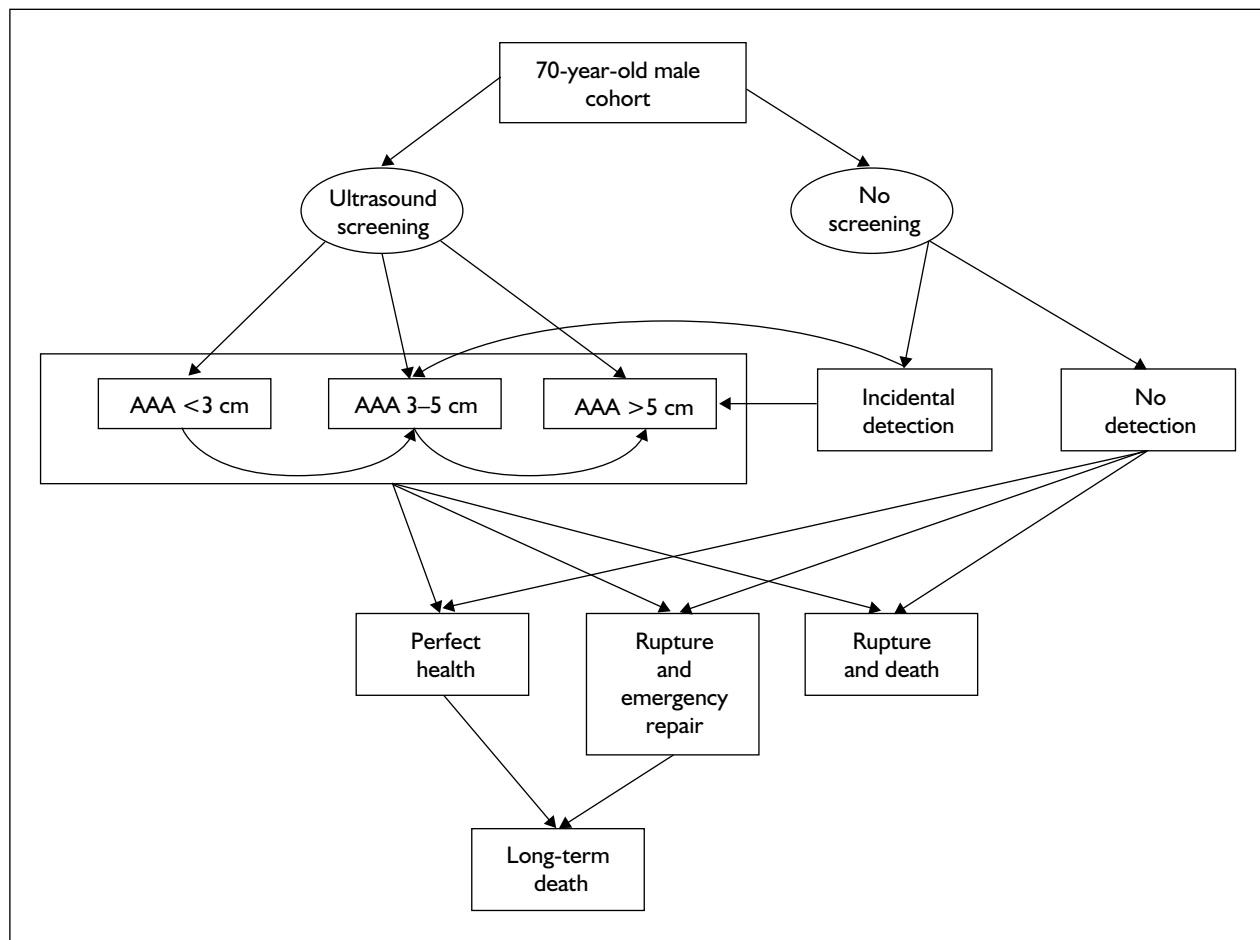
FIGURE 13 *Model structure presented by Lee and colleagues[182]*

would appear to be easier to relax using a patient-level simulation model as the exact time of the next scheduled surveillance screen would be known at the time at which an AAA grew to 5 cm and the appropriate risk of rupture prior to that time could be implemented.

The model structure presented by Lee and colleagues[182] is shown in *Figure 13*, which describes the transition of AAAs from size under 3 cm to between 3 and 5 cm to >5 cm. Separate input parameters were described for AAA sizes 3–4 and 4–5 cm, so the model appears to comprise four AAA size states. It is not clear how the size of AAAs in the <3 cm category were distributed, though these patients had an annual growth rate and a separate risk of rupture.

The presented model structure does not describe the occurrence of complications that were described in the text, which stated that all patients undergoing AAA surgery (elective or emergency) could develop the following complications: renal failure, stroke, MI or major amputation. The only

potentially inappropriate model assumption noted is the application of a constant rupture rate to all AAAs in the largest AAA size category, which may misrepresent rupture rates if such rates continue to be a function of size, as shown in *Table 17*.

Comment on model structure and assumptions

The reviewed models of screening for AAA demonstrated a wide range of model structures and assumptions. The following issues are raised as areas in which alternative approaches were identified in the review, and which may impact on the results of the screening model:

- time horizon
- allowance for screening uptake rates
- appropriate differentiation between AAA sizes
- appropriate cycle length for time-orientated models
- allowance for rupture while under surveillance
- allowance for incidental detection of AAAs
- allowance for contraindication to surgery
- modelling of AAA-related healthcare events.

TABLE 17 *The impact and likely bias of model assumptions made by Lee and colleagues*[182]

Assumption	Impact	Interpretation
AAAs over 5-cm threshold can rupture during surveillance	More realistic representation	No bias
There is a constant rupture rate for all large AAAs	New and existing AAAs in the >5 cm category have the same rupture rate	Likely to favour screening, as ruptures will occur earlier in the model

The use of a limited time horizon is an obvious limitation for any model purporting to measure life-years, or QALYs, as the relevant outcome for an evaluation. It may be argued that a proportion of patients will remain alive forever in a cohort model, but there are commonly adopted techniques to get round this anomaly, such as assuming a maximum age of 100 years, that do not materially affect the results of most evaluations.

Screening uptake rates can affect both the costs and effectiveness of a screening programme as described in more detail in the section 'Progression through undiagnosed cancer states and symptomatic presentation' (p. 42). Costs are affected if there is a large fixed-cost component to the delivery of screening, such that the cost per screen reduces as the number of people screened increases. The effectiveness of screening may vary according to the uptake rate if those who attend and do not attend screening have different prevalence or survival rates. If these factors are incorporated, a model will provide a more accurate representation of screening, but also facilitate analyses of alternative approaches to increase uptake rates. Most of the identified modelling studies did describe a screening uptake rate, although none of them adjusted costs or prevalence by uptake.

Shorter size categories for AAAs are generally preferred in order to differentiate more accurately between alternative screening options. Most of the identified Markov models specified AAA size categories of 10 mm, starting with AAAs sized over 30 mm, although if an individual sampling modelling technique is used, it is possible to describe AAA size as a near continuous variable, for example, specifying size to the nearest millimetre. A reasonable approach is to define the size categories on the basis of the shortest categories for which data informing events related to AAA size are available, such as size-specific growth, rupture and incidental detection rates. It may also be necessary to define size categories beyond the threshold size for surgery. The

advantage of the identified DES model was that it described the continuing growth of AAAs beyond the AAA threshold size for surgery,[172] whereas other models described a single size category above the threshold, e.g. AAAs >50 mm. The application of a single rupture or incidental detection rate to all AAAs in such a category may misrepresent the progression of the AAAs above the threshold size. As time progresses, the distribution of AAAs above the threshold will likely become less right skewed, that is, the mean AAA size will increase. By applying a single rupture rate, the number of ruptures will likely be overestimated as the more appropriate lower rupture rate for smaller AAAs would allow more people to die of competing causes prior to experiencing a rupture.

Related to the choice of size categories is the choice of an appropriate cycle length for a time-orientated model, such as a Markov model. A common statement is that the cycle length should be clinically relevant.[184] In the context of screening for AAAs, clinically relevant may be interpreted as facilitating the progression of AAAs through the defined size categories at an adequate rate; for example, if 10 mm size categories and an annual cycle length are defined, then AAAs in size category 30–40 mm will take a minimum of 2 years to grow to >50 mm.

An apparent shortcoming of most of the identified models is that they did not allow for the rupture of AAAs that were kept under surveillance following detection. There is evidence that AAAs above 30 mm have a risk of rupture. It is clear, therefore, that a proportion of persons with AAAs who are re-screened every 12 months will suffer a rupture in the intervening period. Likewise, there is evidence that a proportion of persons with a detected AAA will be contraindicated to surgery, e.g. the MASS cost-effectiveness study identified around twice as many consultations prior to surgery than episodes of elective surgery.[185] Intuitively, the cost-effectiveness of screening will be influenced by the contraindication rate and this aspect should be modelled.

The final issue relating to model structure and assumptions concerns the modelling of AAA-related healthcare events. As presented in the section 'Presentation of model detail and the relevancy of parameter estimates' (p. 67), there is evidence that persons surviving surgery for AAAs have a shorter life expectancy than the general population. The causes of the shortened survival are the occurrence of other healthcare events that are related to the presence of an AAA, which implies that all persons with an AAA (either detected or not detected, and pre- or post-surgery) are at risk of AAA-related healthcare events. Lee and colleagues[182] modelled four complications post-surgery (renal failure, stroke, MI and major amputation), although they did not allow for the occurrence of similar events pre-surgery, which would likely favour the no screening arm because more persons with AAAs underwent surgery in the screening group. As these events will have both cost and utility implications, a model of screening for AAA should include health states describing related healthcare events for all persons with AAAs.

Choice of modelling technique

The chosen model structure affects the appropriate choice of modelling technique, and the influence of the factors addressed in the previous section on the choice of modelling technique is assessed in this section. Individual sampling models handle the modelling of more complex pathways more easily than cohort models because simultaneous events that affect the progression of patients can be described as attributes that are attached to individuals as they progress through the model. In the case of modelling AAA-related health events, health states may describe progression through a series of AAA size categories, whereas AAA-related events are treated as attributes. Within each size-related health state, each individual may sample a probability of experiencing renal failure, stroke, MI and major amputation as an AAA-related event, and the experience of each event can be stored as an attribute that informs cost and utility weights for that health state, in addition to survival rates.

In a cohort Markov model the tracking of an unruptured AAA at the same time as describing the impact of the related events may cause the exponential expansion of the number of states required by the model. Lee and colleagues[182] described differential annual costs for the first year and subsequent years following an MI, which required two additional states for every AAA size category. *Figure 14* presents a portion of the hypothetical model structure.

Other issues that may make an individual sampling model a more appropriate form of modelling come to light when considering validation sources for the model [as discussed in more detail in the section 'Validation (and calibration)', p. 78]. The MASS trial is the only large-scale UK trial of screening for AAA,[185] and therefore provides a key source for validating the model, which requires setting up a model to reflect the key decision parameters within the trial. The decision parameters include the specification of AAA size categories between 3 and 4.4 cm and between 4.5 and 5.4 cm; 3-monthly screens for AAAs detected in the latter category; and consideration of elective surgery for AAAs over 5.5 cm, with signs of rapid expansion (≥1 cm per year) at a follow-up screen, or showing symptoms attributable to aneurysms. Hence setting up a model that incorporates these decision rules requires the application of alternative surveillance
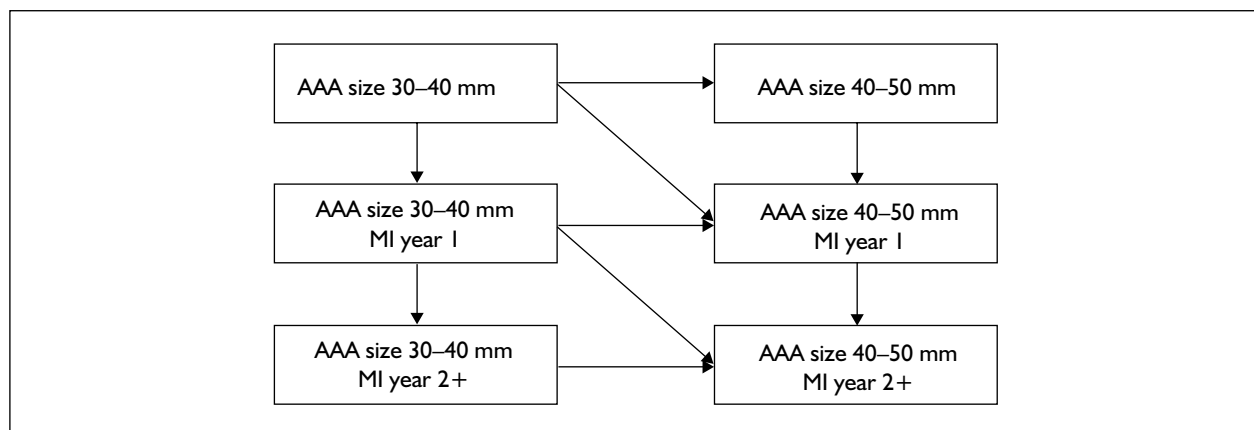

FIGURE 14 *Example of Markov model health state proliferation*

schedules for AAA categories outside the commonly proposed 1 cm categories, and the ability to track fast-growing aneurysms. Both of these factors may be more easily adopted using an individual sampling model.

Presentation of model detail and the relevancy of parameter estimates

It is necessary to present sufficient detail to provide the reader with a clear understanding of the data sources and input parameter point estimates. In addition to describing the extent to which the input data were described, where possible this section also critiques the data sources and data analyses that were used to populate the identified models.

The evaluation reported by Mason[171] was poorly referenced and it is difficult to judge the accuracy of the values for some of the key input parameters. *Tables 18* and *19* describe details of the data sources that were presented and the interpretation of any biases introduced by the described source. In particular, details of the analysis of the referenced data sources informing AAA growth rates and rupture rates would be useful. The model's results were shown to be very sensitive to the relative survival rates post-elective surgery for screened patients and post-rupture for non-screened patients, so more detail on the data source used to estimate the baseline relative survival rates would be useful. The authors stated that the full survival analysis is available, but it can only be presumed by the reader that the survival analysis accounts for the differential timing of events (ruptures and elective surgery) in the

TABLE 18 Uncertainties in the input parameter values used by Mason[171]

Parameter	Reference	Interpretation
Screening uptake rates	Referenced to two relevant studies	Does not affect cost-effectiveness as prevalence and costs are not related to uptake
Screening test characteristics	Not referenced	Favours screening, as 100% sensitivity is assumed
AAA prevalence	Prevalence may be based on AAA mortality and rupture survival rates. Size distribution is not referenced	Unknown bias
AAA growth rates	Not referenced	Unknown bias
AAA rupture rates	Not referenced	Favours screening, as no small AAAs rupture during surveillance
AAA opportunistic detection	Not referenced	Favours screening, as no AAAs are detected opportunistically
Contraindication rate	Not referenced	Unknown bias
Elective mortality rates	Referenced to an editorial	Unknown bias, was the surgical procedure relevant?
Rupture survival rates	Not referenced	Unknown bias, was the surgical procedure relevant?
Differential post-surgery survival	Based on 5-year survival curves, although sample sizes not reported	Unknown bias

TABLE 19 Data sources for cost parameters presented by Mason[171]

Parameter	Reference	Interpretation
Screening costs	Micro-costed, although only summary details provided	Relevant source, could allow definition of fixed and variable costs
Acute surgery costs	Referenced to personal communication	Unknown relevance

screened and non-screened cohorts. It is also necessary to know whether the survival analyses were age-adjusted as patients surviving elective surgery would generally be younger than patients surviving emergency surgery.

The data sources presented by Frame and colleagues[170] are described in *Tables 20* and *21*. Although the data sources are clear, there remains uncertainty around how particular parameter values were estimated, for example, AAA prevalence was referenced to a UK population study that estimated a prevalence of 7.8% and a Swedish study of ultrasonography in hypertensives (an *a priori* high risk group), which estimated a prevalence of only 2%. It is not clear how the aggregate AAA prevalence of 5.4%, or the split between AAAs above and below the 4-cm surgical threshold were defined. Similarly, the assumptions required to estimate aggregate rates of growth for all AAAs under 4 cm to over 4 cm were not defined.

TABLE 20 *Uncertainties in the input parameter values used by Frame and colleagues[170]*

Parameter	Reference	Interpretation
Screening uptake rates	Not mentioned	Does not affect cost-effectiveness, as prevalence and costs are not related to uptake
Screening test characteristics	Referenced, and detailed description of studies provided	Favours screening, as ultrasound sensitivity is assumed to be 100% despite studies indicating rates lower than 100%
AAA prevalence	Referenced to a UK population based study and a Swedish high-risk study	Unknown bias, as analysis and assumptions are not discussed
AAA growth rates	Referenced to 2 observation studies of AAAs *in situ*	Relevant sources, although data analysis is not discussed
AAA rupture rates	Two studies referenced, no details on distribution of AAA size in studies	Likely to favour screening, as constant rupture rate is assumed for all large AAAs
AAA opportunistic detection	Not referenced	Unknown bias, as proportion of AAAs are detected opportunistically
Contraindication rate	Referenced to two studies, including one from the UK	Unknown bias, were the referenced populations relevant to the model population?
Elective mortality rates	Appears to be referenced to two studies from the early 1980s	Unknown bias, was the surgical procedure informing the survival rates relevant?
Rupture survival rates	Appears to be referenced to two studies from the early 1980s	Unknown bias, was the surgical procedure informing the survival rates relevant?
Differential post-surgery survival	Appears to be referenced to two studies from the early 1980s	Likely to favour screening, as post-surgery survival same as normal population

TABLE 21 *Data sources for cost parameters presented by Frame and colleagues[170]*

Parameter	Reference	Interpretation
Screening costs	Medicare reimbursement rates	Relevant source, although recognised as being a high estimate
Acute surgery costs	Referenced to economic study of AAAs	Unknown relevance
Post-surgery costs	A follow-up office visit (US$35) is not referenced	Unknown relevance, as application is not described

Other areas in which additional details on the sources informing parameter values would have been useful include the incidental detection of AAAs, which was not referenced, and survival rates post-rupture and post-elective surgery, which were loosely referenced to two relatively old papers, both of which were only two pages long.

Tables 22 and *23* present the main uncertainties around the data sources used by Law and colleagues.[181] The data sources were generally well referenced and the data analyses clearly described, with most of the relevant data presented in tables. The relevance of some of the data sources may be questioned, particularly the surgical series and the necropsy series that date back to 1964 and 1955, respectively. However, the data are presented so the user can judge the comparability of the earlier and later series.

Life expectancy post-surgery was based on life expectancy in the general population, and incorporated an increased risk of death from other vascular risks in this population. The ratio of deaths from circulatory to non-circulatory causes was based on a cohort of US patients who survived elective AAA surgery and the general US population. The estimated ratio was applied to England and Wales mortality data to estimate mean life expectancy in the UK for men aged 60, 65, 70 and 75 years, though no details of the estimation method were presented. The average number of years of life lost by men dying from a ruptured AAA is presented, presumably as a function of the age distribution of AAA deaths. A small correction may be required if the probability of elective operative death is a function of age.

Costs were attached to the number of elective surgery episodes, emergency surgery episodes and cases of failed hospital resuscitation for persons reaching hospital after rupture but dying prior to surgery. Two ultrasound screens were assumed for each man between the ages of 60 and 80 years, although a follow-up strategy for aneurysms smaller than the threshold is not defined. This may be important as the study assumes that all aneurysms that reach the threshold will be detected, which will be affected by the rescreening strategy.

St Leger and colleagues[174] stated that further details on the "considerations" behind the parameter estimates were available from the authors, which is essential because virtually no details were provided in the paper. The author did respond to an email query about the data sources, although the information had been lost in the intervening period. *Table 24* describes the parameters for which no reference details are provided.

TABLE 22 Uncertainties in the input parameter values used by Law and colleagues[181]

Parameter	Reference	Interpretation
AAA deaths	UK government mortality statistics	Good, contemporary reference
Size of AAA at rupture	Based on 3 surgical series (1964, 1972, 1993) and 2 necropsy series (1955, 1957) (n = 163), smoothed using cubic splines	May favour no screening if surgical studies excluded ruptures that did not make it to surgery (likely to be larger)
Ratio of ruptured to unruptured AAAs at death	Based on 3 necropsy studies (1955, 1957, 1977) (n = 160 for AAAs >7 cm)	Unknown bias, old sources, but sufficient detail presented
Survival	Based on reasonable data sources, but analysis is unclear	Unknown bias, same life expectancies applied to elective and emergency surgery survivors

TABLE 23 Data sources for cost parameters presented by Law and colleagues[181]

Parameter	Reference	Interpretation
Screening costs Acute surgery and failed resuscitation costs	All referenced to single study	Unknown relevance, as details not provided
On average, every man receives two screens	Assumed	Unknown bias, difficult to estimate (and inform policy) without modelling growth rates

TABLE 24 *Uncertainties in the input parameter values used by St Leger and colleagues[174]*

Parameter	Reference	Interpretation
Screening uptake rates	Not referenced	25% screening refusal rate
Screening test characteristics	Not referenced	Favours screening, as assumed to be 100% sensitive
AAA prevalence	Not referenced	Unknown bias
AAA growth rates	Not referenced	Unknown bias
AAA rupture rates	Not referenced	Unknown bias, size-specific rupture rates
AAA opportunistic detection	Not referenced	Favours screening, as assumed to be zero
Contraindication rate	Not referenced	Favours screening, as assumed to be zero
Elective mortality rates	Study from 1986 is referenced	Unknown bias, was the surgical procedure informing the survival rates relevant?
Rupture survival rates	Range of 80–94% is referenced	Favours screening, as 100% mortality assumed
6-cm AAAs at 5 years have 2-year survival	Not referenced	Unknown bias
Differential post-surgery survival	Not referenced	Likely to favour screening, as post-surgery survival same as normal population

An annual risk of rupture and an annual amount of growth were defined for four size categories of aneurysm (3–3.99, 4–4.99, 5–5.99 and >6 cm). The specification of aneurysm growth as *x* cm per year, as a function of the existing size of the aneurysm, resulted in some inaccuracies in the model, for example, aneurysms of size 4.5–4.99 cm had an annual growth rate of 0.46 cm, and were all assumed to progress to the next size category (5–5.39 cm) by the following year. As can be seen, aneurysms of size 4.5 cm would not progress to the subsequent category within 1 year. Such inaccuracies were compounded over the course of the eight aneurysm states.

The estimation of life expectancy after the 5-year rupture model assumed that all men developing an aneurysm ≥6 cm did so at age 70 years. A mean life expectancy of 12 years in the non-affected population aged 70 years was assumed (based on men in England and Wales, although not referenced).

Table 25 describes the assumptions and data sources used to inform the cost and utility parameters within the model. The cost sources appear to be reasonably well sourced, though the screening costs were around 10 years old and were not uprated. The additional survival was assumed to be experienced "at full quality", such that a utility value of 1 was applied, that is, the results are really cost per life-year gained.

Table 26 describes the data sources for the clinical parameters in the model of Pentikainen and colleagues.[172] The probability of a positive screen and the distribution of aneurysm sizes in the targeted population were based on a small empirical study (38 positive cases from 238 screens). Only a few observations of females were available, so prevalence and rupture incidence in females was assumed to be one-quarter of that in males, based on literature estimates.

Some 80% of aortas in the smallest aneurysm category were assumed to be normal and not to grow. All other aneurysms were assumed to grow exponentially at an annual rate of 5%. Two references were quoted in the discussion as "supporting" the exponential growth model, although a separate study found no evidence of increasing absolute growth rates by AAA size,[186] as referenced by Frame and colleagues.[170] Such discrepancies underline the preference for undertaking systematic reviews to inform input parameters. Rupture risk was estimated as an increasing function of size, using a quadratic increasing risk function that gave a zero risk at 4 cm, a 5% annual risk at 5 cm and a 100% risk at 10 cm. The distributional form was assumed, although increasing rupture rates by size were referenced. All persons with a negative screen were assumed not to be at risk of future rupture, and hence died of other causes.

Empirical data were used to estimate separate 15-year survival curves post-surgery for elective and emergency patients. These survival curves were input directly into the model (post-15-year survival is assumed equal to the general

TABLE 25 *Data sources for cost and utility parameters presented by St Leger and colleagues[174]*

Parameter	Reference	Interpretation
Screening costs	A pilot study and a referenced study reporting fixed and variable costs	Useful implementation of fixed and variable costs. Likely to favour screening, as costs have not been uprated from 1986
Surgery costs	An analysis of series of patients treated at a UK hospital	Not clear whether micro-costed, but appears to be a decent source
Utility values	Not referenced	Favours screening, as assumed to be 1 for all health states

TABLE 26 *Uncertainties in the input parameter values used by Pentikainen and colleagues[172]*

Parameter	Reference	Interpretation
Screening test characteristics	Referenced to an editorial from 1985	Favours screening as assumed to be 100% sensitive
Screening uptake rates AAA prevalence	Local empirical study (interpolated numbers: $n = 322$, male screened 132, male AAAs 32, female screened 106, female AAAs 6	Relevant source, but small sample
AAA growth rates	Assumed, although studies giving general support to defined growth rates (although not growth rates at extremes) are referenced	Unknown bias of under- and overestimation of growth rates at extremes
AAA rupture rates	Assumed, although studies giving general support to defined rupture rates are referenced, with apologies for choosing the functional form of the hazard function	Unknown bias, size-specific rupture rates
AAA opportunistic detection	Parameter values not described	Unknown bias
Contraindication rate	Assumed, not referenced	Unknown bias
Elective mortality rates Rupture survival rates	Number of ruptured patients undergoing surgery not referenced. Local empirical data estimated operative mortality (512 elective; 182 emergency operations)	Unknown bias for proportion making it to surgery, otherwise relevant source
Differential post-surgery survival	Survival analysis of 576 elective, and 277 emergency patients surviving 30 days after surgery with maximum 15-year follow-up	Relevant source

population). It was not stated whether the survival analysis accounted for age or whether a constant survival rate was applied to all patients.

Table 27 presents the data sources for the cost parameters in the model (utilities were not measured), which shows that the study used locally relevant empirical data sources.

Soisalon-Soininen and colleagues[173] used the same model as Pentikainen and colleagues,[172] although there appear to be some differences in the parameter estimates between the two models; for

example, Pentikainen and colleagues defined rupture risk as an increasing function of aneurysm size starting from aneurysms of size 4 cm, whereas Soisalon-Soininen and colleagues defined a constant rupture risk for aneurysms larger than 5 cm (smaller aneurysms have a rupture risk of zero). Other differences between the studies included the rupture mortality rate. Both studies estimate that 60% of rupture patients died before surgery, but Pentikainen and colleagues estimated a 38% operative mortality for emergency ruptures, whereas Soisalon-Soininen and colleagues estimated a 64% 30-day operative mortality rate.

TABLE 27 *Data sources for cost parameters presented by Pentikainen and colleagues[172]*

Parameter	Reference	Interpretation
Screening costs	Same empirical study as informed prevalence, details not provided	Relevant source
Acute surgery costs	Same empirical study as informed prevalence (emergency $n = 29$; elective $n = 35$), details provided on resource items included	Relevant source
Postoperative costs	Based on 5-year follow-up data from the Finnish National Discharge Register	Relevant source, although not clear if non-AAA-related resource use is included

The post-surgery survival rates presented by the two studies were identical, so it is not clear why there was a difference in operative mortality. Both studies assumed an incidental detection rate for AAAs in the non-screened population, although the proportion was not stated by Pentikainen and colleagues (Soisalon-Soininen and colleagues reported a 63% incidental finding rate, based on local empirical data).

Lee and colleagues[182] provided virtually no information on the data sources used to estimate input parameter values, and the input parameter values are described as weighted averages. The reader was invited to request references, which is unusual as the presentation of references is not generally subject to space constraints. A letter was sent to the corresponding author requesting further information, but no reply was received. *Table 28* lists the data sources provided, which shows that a number of key parameters were assumed, including contraindication rates and incidental finding rates. In addition to the lack of information on data sources, the methods (and assumptions) for converting the set growth rates into transition probabilities, or for estimating the excess mortality risk for the complications, were not presented.

Table 29 presents the data sources for the cost and utility parameters in the model. The main uncertainties are around the sources for the cost of the complications and the utility values.

Comment on presentation of model detail

The adequacy of the presented information varied from a minimum level at which virtually no references to the data sources were provided,[174,182] to a good standard of referencing and the actual presentation of the raw data derived from the referenced sources.[181] That Law and colleagues presented the most detailed information on the data sources used to model the cost-effectiveness of AAA screening was obviously assisted by the

relative simplicity of the modelling approach adopted, which was based around a limited set of input parameters that excluded unobservable parameters such as AAA growth rates. This approach extends to the estimation of costs where Law and colleagues simply assumed that each man underwent two screens.

The St Leger[174] and Lee[182] studies raise an interesting issue, however, as they invite the user to contact the authors to request additional information about their models, although upon application no further details were forthcoming. The invitation to apply for further details is one solution to journal space constraints, although this approach is only feasible if the authors have a more complete research report that can be dispatched to users. Otherwise, authors are unlikely to be able to respond adequately to enquiries occurring years after the publication of the paper when the authors' recollection of the details will be diminished. A better solution would be for journals to facilitate the publication of additional material as a web-based appendix to the paper-based version. Space constraints will always be an issue for the publication of decision model-based evaluations, given the synthesis of data from a range of sources. Developments in the application of such models, such as the increased frequency of probabilistic sensitivity analyses that incorporate probability distributions around most input parameters, and the fact that good screening models tend to be more complex than treatment models, mean that the space constraints are likely to be felt more acutely by authors of such models. Some journals do provide such facilities (e.g. *JAMA*, *BMJ*), although the uptake of this facility appears to be limited.

Nevertheless, it may be reasonably expected that the paper-based journal version of a modelling study should at least reference the data sources used to estimate the model's input parameters, briefly describe the methods of any data analyses

TABLE 28 *Uncertainties in the input parameter values used by Lee and colleagues[182]*

Parameter	Reference	Interpretation
Screening uptake rates	Assumed, not referenced	Favours screening, as assumed to be 100%
Screening test characteristics	Based on local study of 25 patients, in whom 7 AAAs were detected	Favours screening, as assumed to be 100% sensitive
AAA prevalence	Not referenced, described as a weighted average	Unknown bias, presented by size
AAA growth rates	Presented as a set growth per year (e.g. 1 mm for AAAs less than 3 cm). Not referenced, described as a weighted average	Unknown bias
AAA rupture rates	Not referenced, described as a weighted average	Unknown bias, presented by size
AAA opportunistic detection	Assumed, not referenced	Unknown bias
Contraindication rate	Implied to be zero, not referenced	Favours screening, as most studies assume a positive contraindication rate
Elective mortality rates	Not referenced, described as a weighted average	Unknown bias
Rupture survival rates	Not referenced, described as a weighted average	Unknown bias
Complications rates following elective and emergency surgery	Not referenced, described as a weighted average	Unknown bias
Differential post-surgery survival	Excess mortality rates associated with complications referenced to three studies, but estimation process is not described	Unknown bias

TABLE 29 *Data sources for cost and utility parameters presented by Lee and colleagues[182]*

Parameter	Reference	Interpretation
Screening costs	Derived from the cost accounting system at the authors' US hospital	Relevant source
Acute surgery costs	Referenced to recent economic evaluations	Relevant source
Long-term complications costs	Not referenced, though separate costs are presented for the first year and subsequent years	Unknown bias
Utility values	Not referenced, though method for accounting for the temporary effect of elective surgery is described	Unknown bias

undertaken to estimate important parameters and discuss data limitations for parameters to which the cost-effectiveness results are shown to be particularly sensitive.

Comment on relevancy of parameter estimates

The following sections present brief outlines of relevant estimates for alternative input parameters based on the sources reported in the identified studies. Unfortunately, it was not feasible to undertake a systematic review of parameter estimates in the available time frame.

Screening uptake rates

Most studies compared only one screening option, and none of the studies described screening costs as a function of uptake, so uptake rates affected only total costs and effects, not cost-effectiveness. The most relevant parameter estimates were those based on local screening studies.[171–173]

Screening test characteristics

The only parameter value that did not differ across studies was the sensitivity rate for ultrasound screening for AAAs, which was assumed to be 100%. However, the only study to present a review of the effectiveness of ultrasound as a screening test for AAAs showed that the empirically estimated sensitivity is less than 100%.[170]

Prevalence

Mason[171] appeared to estimate prevalence by dividing observed mortality rates by the rupture

mortality rate to estimate the total number of ruptures. The problem with this estimation procedure is that additional AAAs would be detected by screening that would not rupture prior to death from other causes. Also, Mason did not describe how the size distribution of the prevalent AAAs was estimated.

The only other study to provide any detail on the sources used to estimate AAA prevalence is that of Pentikainen and colleagues,[172] who obtained prevalence estimates by size of AAA from a relatively small screening study in which 38 AAAs were detected from a population of 238 high-risk individuals. Other studies reference prevalence estimates, although even after reviewing the referenced papers it is sometimes not clear how the referenced studies were used to inform prevalence rates.

Given the high sensitivity of the ultrasound screening test, screening pilot studies or screening trials are likely to provide fairly accurate estimates of age-specific AAA prevalence by size, although there may be concerns over the representativeness of attenders. However, in the context of a screening evaluation, the prevalence of AAA in the population likely to be screened may be more relevant than the full population prevalence as the non-screened population will have the same costs and benefits in the presence and absence of a screening programme. Ideally, prevalence rates in both populations would be used to inform the evaluation of increased screening uptake.

It is also important that prevalence is estimated across ages that are potentially eligible for screening. A brief review of the literature identified few relevant sources in the area of AAAs, although the Collaborative Aneurysm Screening Study Group[187] identified a highly significant difference in prevalence rates between screening studies undertaken in the UK, Australia and Denmark, suggesting that locally derived data are important for informing prevalence. Age-specific prevalence rates were identified in a Danish population-based study,[188] though the rates corresponding to the MASS trial were much higher than observed in the UK study (12.8% versus 5% prevalence for AAAs larger than 3 cm). Lederle and colleagues[189] showed a highly significant difference in prevalence rates between persons who have ever smoked and those who have never smoked. The MASS trial report does not present the proportion of smokers in the screened population,[185] so this potential difference cannot be assessed. In the absence of a systematic

TABLE 30 *Adjusted age-specific prevalence estimates*

Age (years)	Danish data		Adjusted data	
	>29 mm	>39 mm	>29 mm	>39 mm
25–44	0.000	0.000	0.000	0.000
45–54	0.019	0.000	0.007	0.000
55–64	0.06	0.011	0.023	0.004
65–74	0.128	0.041	0.050	0.016
75–84	0.185	0.086	0.072	0.033
Total	0.082	0.023	0.032	0.009

review of the literature, and access to patient-level data from relevant studies, the age-specific Danish prevalence figures could be adjusted downwards by applying a constant ratio of the prevalence of AAAs in the age range 65–74 years observed in the Danish and the MASS trial to the observed rates in all age groups of the Danish study, as presented in *Table 30*.

Other sources that are commonly used in cancer screening models are necropsy studies; although based on the sources presented by Law and colleagues,[181] there are unlikely to be any recent such studies, which reduces their relevance due to possible birth cohort effects.

Incidence
Frame and colleagues assumed an incidence rate of new AAAs of 0.1% as no direct data informing incidence rates were identified;[170] no other studies modelled incidence. Given the almost perfect sensitivity assumed for ultrasonography, it follows that repeat screening can be excluded as an option if the incidence rate is so close to zero. However, a more recent study found a 4-year incidence rate of 2.2% (95% CI 1.6 to 2.8%) in over 2300 men (mean age 66 years at second screen) who were rescreened 4 years after an initial screen that had shown them to be clear of AAA.[183] In addition, there was a 0.4% (20/5151) rate of interim AAAs, of which only one showed strong signs of being a false-negative screening result. Lederle and colleagues concluded that a second screen after an interval of more than 8 years may provide similar yields to the prevalence screening round.[189] This indicates that screening models should evaluate the cost-effectiveness of alternative repeat screening strategies for AAAs.

Growth rates
Several studies describing growth rates of AAAs were referenced, although no details of the primary data were provided. Pentikainen and colleagues[172] developed growth and rupture risk

equations that were "supported" by the literature, although they were not validated in the model.

Given the intervention strategy for detected AAAs, that is, that surgery may be delayed until the AAA reaches a size of 6 cm in some cases, and that there is a sizeable proportion of patients in whom surgery is contraindicated, there are likely to be a number of studies reporting growth rates for AAAs. Indeed, a simple MEDLINE search combining the MeSH term 'Aortic Aneurysm, Abdominal' with the freetext term 'growth rates' identified four relevant studies published since between 1996 and September 2004.[190–193]

Vardulaki and colleagues,[190] in particular, presented separate estimated exponential growth rates for AAAs at three initial diameters (3, 4 and 5 cm) that were shown to fit the presented screening data closely, and also histograms of the observed growth rates that show the distributions to be approximately normally distributed. A growth rate of 1.02 could be converted to transition probabilities from one AAA size to another, say 30–40 mm, using the following approach:

$$\text{size}_{t1} = \text{size}_{t0}(\text{growth}^n) \rightarrow 40 = 30(1.02^n)$$

where n is the number of years required for the mean AAA to grow to size 40 mm. As when $a^n = x$ then $n = \log_a x$, then $n = \log(40/30)/\log(1.02) =$

14.53 years, which can be converted to an annual transition probability as $1 - \exp(-1/14.53) = 0.067$.

Similar estimates can be undertaken for alternative starting AAA sizes, although this approach requires consideration of two particular issues. First, if size categories are used to describe AAA size, then alternative model 'start' and 'within model' states may be required for each category, for example, for a state '40–50 mm AAAs' it may be assumed that the mean size of AAAs in this state at the start of the model is 45 mm. AAAs entering this state once the model starts will do so at size 40 mm. Therefore, alternative transition probabilities should be applied to these two groups of AAAs.

No data describing growth rates of AAAs above starting size 50 mm were identified, but the observed increase in growth rates in 30-, 40- and 50-mm AAAs may be used to estimate growth rates at larger starting sizes. *Table 31* presents the set of transition probabilities assuming that the AAA growth rate increases at a constant rate, as observed between sizes 40 and 50 mm (2.9%). *Table 32* presents the estimated transition probabilities if it is assumed that growth rates increase at a growing rate, as informed by the ratio of the growth rates between 40 and 50 mm to the rates between 30 and 40 mm (2.11).

TABLE 31 *Estimated annual transition probabilities between alternative size categories of AAA assuming a constant increase in the AAA growth rate*

Start size (mm)	30	40	50	60	70	80	90
End size (mm)	40	50	60	70	80	90	100
Annual growth rate (mm)	1.02	1.03	1.06	1.09	1.13	1.16	1.19
Log(end size/start size)	0.12	0.10	0.08	0.07	0.06	0.05	0.05
Log(growth rate)	0.01	0.01	0.03	0.04	0.05	0.06	0.08
$N = \log(\text{end size/start size})/\log(\text{growth rate})$	14.53	6.67	2.94	1.70	1.12	0.80	0.60
Transition probability	0.07	0.14	0.29	0.44	0.59	0.71	0.81

TABLE 32 *Estimated annual transition probabilities between alternative size categories of AAA assuming an increasing increase in the AAA growth rate*

Start size (mm)	30	40	50	60	70	80	90
End size (mm)	40	50	60	70	80	90	100
Annual growth rate (mm)	1.020	1.034	1.064	1.129	1.276	1.625	2.567
Log(end size/start size)	0.12	0.10	0.08	0.07	0.06	0.05	0.05
Log(growth rate)	0.01	0.01	0.03	0.05	0.11	0.21	0.41
$N = \log(\text{end size/start size})/\log(\text{growth rate})$	14.53	6.67	2.94	1.27	0.55	0.24	0.11
Transition probability	0.07	0.14	0.29	0.55	0.84	0.98	1.00

The observed differences according to the alternative assumptions could have an important impact on the estimated cost-effectiveness of screening. Thus, a systematic review and meta-analysis would ideally be required to populate these parameters, particularly as Frame and colleagues[170] referenced a paper that found no evidence of an increasing growth rate, let alone an increasing increase in the growth rate

The development of theories to inform initial estimates of such parameters, which then form the basis for the calibration of the model against external data sources, provides a more robust approach to populating screening models. None of the identified studies calibrated their models to estimate unobservable parameters, such as AAA growth rates. The process of calibration is discussed in more detail in the section 'Model calibration' (p. 38). In an AAA screening model, the assumed constant increase in growth rates could be varied during a process of calibration. Potential sources for calibrating or validating a model of screening for AAA are discussed in the section 'Validation (and calibration)' (p. 78).

Alternatively, data synthesis techniques such as the Markov chain estimation procedure proposed for estimating non-observable transition probabilities in the colorectal cancer screening case study, which use primary data describing the incidence of alternative stages of the disease (see the section 'Progression through undiagnosed cancer states and symptomatic presentation', p. 31), could be applied to similar data sources concerning AAAs.

Rupture rates
Law and colleagues[181] demonstrated a direct method of estimation for size-specific rupture rates based on a series of longitudinal studies reporting rupture rates by starting size of AAA. A logistic regression model was fitted to these data to estimate the annual risk of rupture; Vardulaki and colleagues[190] stated that the analysis presented by Law and colleagues provides the best estimates of rupture rates.

AAA opportunistic detection
None of the identified modelling studies referenced a data source for the estimation of a non-screen detection rate of AAAs. This is a particularly difficult parameter to estimate, being ideally informed by screening trials in which the number of opportunistically detected cases in the non-screening group can be compared with the total number of detected cases in the group randomised to screening. The proportion of AAAs detected opportunistically (assuming 100% sensitivity of the screening test) would be estimated as the additional cases detected opportunistically in the no screening group divided by the number of cases detected by screening in the screening group.

There are complicating issues regarding patients identified as having aneurysms outside of the screening programme, and where they lie in relation to the screened population. Taking the figures from the MASS trial,[185] a total of 70,495 men were assessed for eligibility, of which 67,800 were randomised. Ninety-two elective operations were recorded in the control group of approximately 34,000 (0.27%); 27,000 of the 34,000 randomised to the screen arm of the trial were actually screening, and the total number of aneurysms identified requiring surgery was 322 (291 within and 31 outside protocol). If it is assumed that the 31 surgery episodes identified outside protocol were in the non-attenders, the rate of surgery in the screened group was 1.08%. The opportunistic identification of patients requiring surgery would be estimated at approximately 0.27/1.08 = 0.25.

This is a very approximate calculation and further modifications may be required to address two further issues. First, incidence in the screened group may be different to incidence in the control group. Second, the opportunistic identification of aneurysms within the trial may not be equivalent to that elsewhere, and it may be that the heightened awareness of aneurysms in the geographical area concerned in the trial may mean that this was inflated compared with other areas.

Contraindication rates
Frame and colleagues[170] referenced two papers that appeared to present different estimates of the contraindication rate (10–40%), and so took a rate of 30% as the baseline estimate. No other papers referenced studies reporting contraindication rates.

Elective and emergency operative mortality rates
A range of studies reporting operative mortality rates (in addition to preoperative mortality rates for ruptures) were referenced, although all sources dated back to the 1980s. Pentikainen and colleagues[172] based their estimates on a small local screening study. Alternatively, a study by Michaels[194] was published after all of the identified models, although it was based on readily available data (hospital episode statistics) that

provide as good a source as one might expect to find on such parameters in the UK. A total of 38,319 abdominal aortic procedures or deaths in hospital with a primary diagnosis of aortic aneurysm were identified. The elective mortality rate was 6.4% overall; postemergency operative mortality was 35%, rising to 41.6% if urgent procedures were excluded and 63.1% if non-operated cases were included.

Postoperative survival

Four broad approaches to modelling postoperative survival have been used. First, Frame and colleagues[170] and St Leger and colleagues[174] both assumed that survivors had the same survival profile as the general population. Second, Law and colleagues[181] assumed that all operative survivors had the same survival rate, but one that was reduced compared with the general population. Third, Mason[171] and Pentikainen and colleagues[172] estimated separate survival curves for elective survivors and for emergency operative survivors. Finally, Lee and colleagues[182] modelled alternative complication rates following both forms of surgery, although it is implied that age-specific annual mortality rates were applied to all health states.

From the evidence presented, the use of general population mortality would appear to be a questionable assumption, as the AAA population appears to be at increased risk of other cardiovascular events. A persuasive argument is that only the strongest survive a rupture, and so are likely to have longer mean survival than the broader collection of elective surgery survivors, as shown by Mason.[171] However, Soisalon-Soininen and colleagues[173] presented separate survival curves that showed elective patients to have better postoperative survival rates, which may be because the long-term effects of the rupture are more severe, or because elective surgery includes a proportion of patients whose AAA would not have ruptured before death from other causes, who may be healthier than the remainder of the AAA population. In the absence of a systematic review of postoperative survival, an assumption of equally reduced survival compared with the general population for postelective and postemergency surgical survivors would seem to be reasonable.

As noted at the end of the section 'Model structure and assumptions' (p. 59), all persons with an AAA may be at increased risk of AAA-related events that have cost and utility effects that should be included in an AAA screening model because there will be different numbers of such events in

the screening and no screening groups. Lee and colleagues[182] are the only authors who modelled AAA-related events, although they did not model pre-surgery event rates and did not provide any details about the data sources used to inform their specified event rates. Modelling AAA-related healthcare events involves the specification of the range of related events to be modelled, the estimation of their incidence rates, and their subsequent mortality rates. This is a substantial task, involving consultations with clinical experts and the estimation of parameters for which relevant data are scarce, which may explain the relative absence of such modelling efforts in the identified modelling studies.

Cost parameters

A range of sources and methods were presented for the estimation of the cost parameters, including published charges,[170,182] references to other economic studies and micro-costed estimates of the costs of screening.[171,174] The two examples of micro-costing are preferable as they allow the user to assess directly the relevance of the parameter estimates. They also allow the application of fixed and variable costs that could be used to adjust costs to different uptake rates.

St Leger and colleagues[174] and Pentikainen and colleagues[172] presented elective costs based on series of elective (23 and 35 patients, respectively) and emergency (eight and 29 patients, respectively) operations. Pentikainen and colleagues described the range of resource use items included in the analysis, unlike St Leger and colleagues. The remaining studies referenced other studies for surgical costs, with no details of the reference source provided. Law and colleagues[181] are the only authors to include a cost of failed resuscitation for patients dying before surgery post-rupture, details of which were not provided.

Pentikainen and colleagues[172] used data on 596 patients obtained from the Finnish National Discharge register based on the diagnosis-related group (DRG) codes for all admissions in the first 5 years following AAA surgery to describe postoperative costs (no differentiation is made between elective and emergency surgery survivors). The Finnish data source is useful, particularly as it allows episodes related to the experience of AAA to be separated from other non-related episodes (if required).

Recent guidance from NICE on methods for economic analyses states that all healthcare costs related to the primary condition should be

included. There are a wide range of conditions that could be defined as being related to AAA, which makes the modelling of the downstream costs difficult. In the absence of long-term patient-level data, which would allow a consensus-based approach to the inclusion of the observed healthcare episodes, it is necessary to model the experience of a set of health states that represent the cost (and utility) implications of elective and emergency surgery for AAA. Only Lee and colleagues[182] model differential post-surgery states that allow the estimation of costs relating to renal failure, stroke, MI and major amputation.

From a UK perspective, the recent publication of the cost-effectiveness analysis undertaken alongside the MASS trial provides relevant cost estimates that can be used in an AAA screening model.[195] The MASS group estimated separate costs for the invitation and reinvitation to attend screening, an initial screen and a recall screen. Although only aggregate costs per event are presented in the published paper, access to the data would enable the estimation of fixed and variable costs. A similarly detailed costing exercise was undertaken to estimate the costs of consultation before elective surgery, elective surgery and emergency surgery. Hospital admissions related to AAA within 12 months of surgery were included in the surgical costs, but no other post-surgery costs were included.

Ideally, a model should include similar states for persons with AAAs (either detected or not) as all such persons are likely to be at increased risk of AAA-related events.

Utility parameters

Only Lee and colleagues[182] reported a true incremental cost per QALY gained of screening, which applied utility values <1 to all of the post-surgery complication states, in addition to assuming alternative numbers of days of 'disuse' following elective and emergency surgery. No references to the utility assumptions were presented.

As alluded to in the previous section, a potential problem with this approach is that all persons with AAAs pre-surgery are assumed to have a utility of one, but some of the complication states (e.g. MI and stroke) will be related to the presence of an AAA rather than the surgical process. Therefore, the cost and utility effects of such complications should also be modelled in the pre-surgery AAA population, otherwise the impact of the larger proportion of pre-surgery AAAs in the non-screened group will be underestimated.

The only other identified utility data were collected as part of the MASS trial, which presented utility values for persons experiencing negative and positive screening results 6 weeks after screening and 3 and 12 months post-surgery, and from a surveillance group of persons without an AAA (for comparison with the surgery group). As for the extrapolation of costs, in the absence of longer-term patient-level utility data, a screening model should describe healthcare events related to AAA that affect utility, such that appropriate utility values can be attached to these states.

Validation (and calibration)

The general process of validation or calibration is discussed in the section 'Model calibration' (p. 38). Validation and calibration are similar processes. Validation involves fully populating a model with informed values for all input parameters and then comparing the outputs from the model with relevant observations of the same outputs. If there are differences between the predicted and observed outputs, the analyst should identify likely causes, verify that the differences are in the direction expected, reassess the relevance of the model's data sources, and if necessary, seek further data (perhaps in the form of expert elicitation).

Calibration involves populating the model's input parameters for which relevant data have been identified, and then fitting the remaining parameter values to any observable data that describe model outputs. The wider the range of observable data the more accurate will be the validation or calibration of a model.

Surprisingly, none of the identified studies mention validation, nor do any of the studies appear to calibrate their models to populate unobservable parameters. Law and colleagues[181] did not validate their model, although the model was based around age-specific AAA mortality rates, which implicitly validates the non-screening cohort. Given the weak assumptions around compliance and contraindication to surgery rates, it is unlikely that the screening cohort would be validated against an observed screened cohort.

The main criterion for data sources to inform validation or calibration is that the data source has not been used to populate the model. In the case of AAAs, where population screening is not offered outside clinical trials, nationally collected data sources can be used to validate the non-screened

arm of a screening model. Relevant sources include data describing deaths from ruptured AAA by age, which are published by the ONS, and also health service data describing the number of elective and emergency surgery episodes. Other survey data may also be available to inform the incidence of ruptures. It may be important to validate separately the predicted number of ruptures as the predicted number of deaths may be overly influenced by the chosen rupture mortality rate.

A key data source for informing the accuracy of the screened cohort is the MASS trial,[185] although as described in the section 'Presentation of model detail and the relevancy of parameter estimates' (p. 67), this source may also inform a range of input parameters. However, the validation of a model's outputs with the trial results, based on the same decision parameters as used in the trial, will still provide some reassurance as key parameters, such as growth rates, would be informed by other sources.

In the context of validating an AAA screening model to a screening data source, it is particularly important that the model matches all relevant decision parameters, which may require alterations to the size categorisation of AAAs. In the MASS trial,[185] for example, aortas less than 3 cm were classed as normal and received no further follow-up. Aortas between 3 and 4.4 cm were allocated to annual hospital screens and aortas between 4.5 and 5.4 cm received 3-monthly screens. Aortas over 5.5 cm, with signs of rapid expansion ($\geqslant$1 cm per year) at a follow-up screen, or showing symptoms attributable to aneurysms, were assessed for elective surgery.

Cross-validation: comparison of alternative evaluations

All evaluations, and model-based evaluations in particular, should provide a comparison of their results with the results of previous evaluations. Such comparisons provide a further method of validation, where similar results derived from different evaluative processes provide additional assurances that the observed results are correct. Alternatively, if dissimilar results are observed, it is necessary to identify possible causes for the divergence, which may be due to structural assumptions about the disease process or to the values assumed for particular parameters.

Of the identified evaluations of AAA screening, St Leger and colleagues[174] compared their results with two previously published UK-based models.[171,181] Their conclusions regarding the

appropriate screening policy – screening with an elective surgical threshold of 6 cm – were cited as being in agreement with the results of these evaluations, although the specific results do differ; for example, St Leger and colleagues estimate a cost per QALY of £20,700 for screening with a surgical threshold of 5 cm, whereas Mason finds screening to be dominated for the same strategy.

Soisalon-Soininen and colleagues[173] state only that previous evaluations did not include health care costs after AAA surgery, although no comparisons of results were presented.

Lee and colleagues[182] presented a table summarising five previous evaluations, including a dichotomous category as to whether screening was found to be cost-effective, and the following details of each study:

- study location and date of publication
- eligible population
- surgical threshold size of AAA
- study time horizon
- Markov model or not
- lifetime costs and outcomes considered
- incidental detection considered
- prevalence, risk of rupture, and growth rates of AAA considered.

The text referred to the presented table, although no additional discussion of the differences between the studies is given.

The following section includes a comparative analysis of the results presented by the identified studies, which illustrates how such analyses can inform the model user.

Presentation of model results

This section describes the baseline results presented by the identified models, and also a summary of any sensitivity analyses undertaken. *Tables 34–36* in Appendix 2 summarise the decision parameter values, structural assumptions and input parameter values chosen by each study, which facilitate a comparative analysis of the presented results.

Mason[171] estimated that screening actually reduces the average number of life-years gained, in addition to incurring extra costs, that is, screening is dominated. A range of one-way sensitivity analyses indicated that the results were extremely sensitive to certain parameters, including the annual rupture rate and the differential adjustments to post-surgery survival.

Frame and colleagues[170] presented separate incremental results for ultrasound and physical examination screening for one-off screening, and two-time screening with a follow-up screen 5 years after the first screen. One-off ultrasound screening was found to be moderately cost-effective, with a cost per life-year gained of US$45,550. Only best case [incremental cost-effectiveness ratio (ICER) US$5389] and worst case (screening dominated) sensitivity analyses were presented, so it is not clear which were the most important parameters affecting cost-effectiveness. The possibility of assuming alternative prevalence rates (of small and large aneurysms) in different age groups was discussed and an illustrative example provided.

The baseline results presented by Law and colleagues[181] showed a cost per life-year saved of £746. No sensitivity analyses were presented, although there was some discussion over further data that would usefully inform the precise configuration of a screening programme (it is stated that screening should definitely be provided). Such data included the risk of rupture to inform better the threshold for the offer of surgery, aneurysm growth rates to inform screening intervals, treatments to reduce aneurysm growth rates and risk indicators for surgical mortality.

St Leger and colleagues[174] showed screening to be extremely cost-effective (£1206 per QALY gained). A range of one-way sensitivity analyses showed that the key parameter was the annual risk of rupture. Analysis of the offer of elective surgery to men with aneurysms ≥5 cm showed a marked increase in the cost per QALY gained (£21,000–30,000).

The baseline results presented by Pentikainen and colleagues[172] showed that the cost per life-year gained from screening men was around £5500 and for women £6250. CIs were presented around the mean estimates, but these appeared to represent first-order uncertainty or simple risk (uncertainty about outcomes), rather than second-order uncertainty (uncertainty about probability distributions of outcomes), which is of limited relevance to resource allocation.[196] A range of one-way sensitivity analyses showed that the most important parameters were the annual growth rate of aneurysms, preoperative rupture mortality and the annual rupture risk.

The differences between the studies of Pentikainen and colleagues[172] and the Soisalon-Soininen and colleagues,[173] noted in the section 'Presentation of model detail and the relevancy of parameter estimates' (p. 67), led to a difference in the baseline results for screening first-degree male relatives, with the latter estimating a lower cost per life-year gained of around £3600. A range of one-way sensitivity analyses showed the results to be stable. No extreme values were tested, though the parameter estimates were stated to be conservative (i.e. not favouring screening).

In the base-case analysis presented by Lee and colleagues,[182] screening for AAA was cost-effective with a cost per QALY of US$11,215. Reducing the cost of screening from $259 (approximate Medicare reimbursement) to $40 (the quick screen) reduced the ICER to $6850, and screening high-risk populations improved the ICER to $8460. A threshold analysis on the rate of incidental detection found that the rate would need to be 36% to breach a cost-effectiveness threshold of $60,000 per QALY.

The baseline results of the MASS trial cost-effectiveness analysis,[195] over a 4-year time horizon, was £28,400 per life-year gained, with a 95% CI of £15,000 to £146,000. When survival gain is based on all-cause mortality, the ICER reduces to £13,393. The projected cost-effectiveness over a 10-year time horizon is around £8000.

The above results show that Law and colleagues[181] were the most optimistic with respect to the estimated cost-effectiveness of screening for AAA, with the results of St Leger and colleagues[174] in the same region. Both studies modelled a 6-cm surgery threshold, and assumed that no patients were contraindicated to elective surgery and that there was no incidental detection in the absence of screening. Both studies estimated similarly low costs relative to the other studies, particularly for the costs of screening. Differences include that Law and colleagues assumed that 18% of patients survive rupture compared with a zero survival rate, and that St Leger and colleagues assumed normal life expectancy for elective surgery survivors, while Law and colleagues describe a similarly reduced life expectancy for all surgery survivors. The latter assumptions made by St Leger and colleagues were the more favourable to screening, which suggests that their limited time horizon (5 years) and the combination of the growth rates and rupture rates (which are implicit in Law and colleagues' approach) were sufficiently detrimental to screening to reverse the results comparison.

The ICERs presented by St Leger and colleagues for a 5-cm surgery threshold were significantly

higher than other studies using a similar threshold, such as the analyses presented by Pentikainen and colleagues[172,173] and Lee and colleagues,[182] which showed screening to be less cost-effective, but still well within normally accepted boundaries. The latter studies used lifetime models, which is the most likely explanation of the discordant results. There was no discernible difference in rupture rates, and it is not possible to compare post-surgery survival as Pentikainen and colleagues do not present the data. There were considerable differences between the two models in other areas. The values estimated by Pentikainen and colleagues appeared to favour screening for prevalence, and the rupture mortality rate, whereas the values used by Lee and colleagues favoured screening for growth rates, the incidental detection rate and the elective contraindication rate. The costs presented by Lee and colleagues were higher than those associated with the Pentikainen study (for any feasible exchange rate between US dollars and UK sterling), although the relative costs for elective and emergency surgery were similar to other studies. The slightly lower relative cost of elective surgery in the two US studies is unlikely to compensate for the much higher assumed cost of screening.

Frame and colleagues[170] estimated a higher ICER based on a surgical threshold of 4 cm, which was chosen on the basis of available data rather than clinical reality. In addition, they estimated lower prevalence rates and relatively high incidental detection rates, contraindication rates and elective mortality rates. However, they also assumed no interval ruptures for AAAs detected below the surgery threshold, a high rupture mortality rate and normal life expectancy post-surgery.

Mason[171] is the only author to find that screening was dominated by no screening, that is, costs more and produces fewer life-years. Moreover, certain model assumptions favour screening, such as a zero interval rupture rate for detected AAAs, high rupture mortality rates and no incidental detection in the non-screened cohort. Other parameters are less favourable, including the low prevalence rate, elective surgery contraindication and mortality rates, and a more favourable survival profile for emergency surgery survivors. Mason presents a total cost of screening, although it is not possible to estimate a cost per screen from the data presented. The estimated 1:0.83 ratio between emergency and elective surgery costs is the highest observed, and will favour the no screening option.

The comparison with other models should inform both the sensitivity analyses and discussion accompanying the presentation of a model, which should convince the user that the presented model is the most relevant model to address the defined policy question. This requires an assessment of the relevance of alternative approaches, and where the relative relevance of the adopted approach cannot be definitively stated, sensitivity analyses that test the impact of alternative approaches should be presented. This process is dependent on the explicit presentation of data sources and methods of parameter estimation by previous studies that allow a full assessment of the relevance of other approaches.

Discussion

The focus of this chapter has been on the users' perspective of models for evaluating screening programmes, which has reviewed seven identified cost-effectiveness models of screening for AAAs with respect to four broad issues that affect the usefulness of the model from the users' perspective:

- the relevance of the policy question(s) addressed
- the model structure and assumptions
- the choice of data sources and the presentation of sufficient information for the results to be interpreted in the context of alternative decision areas
- validation, including cross-validation with other models of screening for AAA.

The review found that most studies did not consider more than one main screening option, and the range of separate policies evaluated by individual studies indicates that there are numerous policy options that could have been assessed by each model. Three key policy parameters appear to be the eligible screening population, one-time screening or repeat screening and the AAA size threshold for surgery.

The majority of studies defined the eligible population as all men at a specified age (e.g. 70 years old), or between certain ages (e.g. 60–80 years old). An appropriate policy question should first define the optimal age at which the prevalence screen should be offered when the programme is operating in equilibrium. Second, the maximum age to which it is cost-effective to offer a prevalence screen at the point of implementation of a screening programme could also be determined.

It may also be feasible to evaluate targeted screening of high-risk groups. The broadest approach would be to specify all relevant combinations of risk characteristics by which an eligible population could be specified, for example, patients with hypertension, patients with hyperlipidaemia, patients with hypertension and hyperlipidaemia. Population-based screening should generally be included as policy makers may need to provide an economic justification for excluding the most equitable policy option.

Only one study explicitly modelled the cost-effectiveness of repeat screening.[170] The exclusion of this option may be justified if one-off (or prevalence) screening is not found to be cost-effective, as incidence screening will likely be less cost-effective. However, most studies found one-off screening to be cost-effective, but still did not evaluate repeat screening. Few data describing incidence rates may have been available at the time of analysis of most studies, but further discussion justifying the exclusion of repeat screening as a policy option would better inform the user. Preferably, the model would evaluate repeat screening at alternative intervals.

There does not appear to be a consensus regarding the appropriate AAA size threshold above which elective surgery is offered, which varies from 4 to 6 cm. It therefore appears most sensible to test the cost-effectiveness of AAA screening given alternative treatment thresholds. There was also some variation in the size at which an AAA qualified for further surveillance, including 2, 3 and <4 cm. These options, too, could be investigated.

The above comments illustrate the range of potential policy options that could be evaluated by a good AAA natural history model that adequately addresses the following issues:

- time horizon
- allowance for screening uptake rates
- appropriate differentiation between AAA sizes
- appropriate cycle length for time orientated models
- allowance for rupture whilst under surveillance
- allowance for incidental detection of AAAs
- allowance for contraindication to surgery
- modelling of AAA-related healthcare events.

Any model estimating life-years or QALYs should cover the lifetime of the eligible population. Ideally, screening uptake rates should be linked to both the costs and the prevalence of AAAs in those

attending and not attending screening, although no linkages were identified in the reviewed models.

Shorter size categories for AAAs are generally preferred in order to differentiate more accurately between alternative screening options, although a reasonable approach is to define the size categories on the basis of the shortest categories for which data informing events related to AAA size are available, for example, size-specific growth, rupture and incidental detection rates. Unlike many of the reviewed studies, it is important to continue the categorisation of AAA sizes beyond the surgery threshold size to reflect the changing distribution of AAA sizes above the threshold. If no data are available for one or more event type and a calibration approach is pursued, then a trade-off between the accurate interaction between the size categories and related events and the number of parameters to be fitted must be reached.

Related to the choice of size categories is the choice of an appropriate cycle length for a time orientated model, such as a Markov model, which should facilitate the progression of AAAs through the defined size categories at an adequate rate; for example, if 10-mm size categories and an annual cycle length are defined, then AAAs in size category 30–40 mm will take a minimum of 2 years to grow to >50 mm.

The rupture of AAAs under surveillance following detection, contraindication rates for elective surgery and the incidental detection of AAAs in the absence of an organised screening programme have an intuitive impact on the cost-effectiveness of screening and should be modelled. Another factor that has been shown to be a key uncertain parameter is the description of post-AAA survival, where there is evidence that such individuals are at increased risk of AAA-related healthcare events that influence survival. There are cost and utility implications associated with the causes of the shortened survival, such that a model of screening for AAA should include health states describing related healthcare events for all persons with AAAs.

The above issues affect the appropriate choice of modelling technique for the evaluation of screening for AAAs. Individual sampling models handle the modelling of more complex pathways more easily than cohort models because simultaneous events that affect the progression of patients can be described as attributes that are

attached to individuals as they progress through the model. As the model running time for an individual sampling model will likely be significantly longer than for a cohort model, the analyst should carefully consider the additional benefits of the more flexible approach that the individual sampling approach generally facilitates.

This chapter has also reviewed data sources to inform different categories of parameters that may be specified in a model of screening for AAA. Data sources to inform almost all aspects of an AAA screening model were identified. The primary missing data concerned AAA growth rates beyond 5 cm. The other main area of absent data concerned the modelling of AAA-related healthcare events, which were addressed by only one study. As Lee and colleagues[182] did not describe any of the data sources used to inform this aspect of the model, it is not possible to critique their approach and the timescale of this review precluded further investigation of this aspect of the modelling process.

None of the reviewed studies undertook any form of validation or calibration. Validation is an important process for any model, although it may be considered even more important for screening models due to the role of unobservable parameters. Although potential data sources were identified for most aspects of an AAA screening model, growth rates beyond 5 cm remains an important omission. Relevant sources include data describing deaths from ruptured AAA by age, which are published by the ONS, and health service data describing the number of elective and emergency surgery episodes.

An alternative validation process is the cross-validation with previous models. The strength of a model will be increased if any variations in the model's results can be explained by differences in the assumptions and input parameter values used by alternative models.

Chapter 7
A review of models evaluating antenatal screening programmes

Introduction

Models evaluating antenatal screening programmes generally cover a shorter time horizon than models of adult screening programmes. This chapter reviews identified screening models that have evaluated antenatal screening programmes for three diseases. Risk of autosomal recessive disorders is based on the carrier status of parents, risk of chromosomal abnormalities may be based on a combination of characteristics of the mother and unborn child, and screening for infectious diseases involves the identification of the condition of interest in the mother.

The aim of antenatal screening for autosomal recessive disorders and chromosomal abnormalities is to obtain a measure of the risk that the unborn child is affected by the condition of interest. If the risk is estimated to be above a certain level, the couple (or mother alone) are presented with the estimated risk of an affected pregnancy, and offered the option of a prenatal diagnosis (PND). PND requires the analysis of amniotic fluid, either through the use of chorionic villus sampling (CVS) or amniocentesis, both of which procedures carry a risk of procedure-related spontaneous abortion. If the offer of PND is accepted, and the presence of an affected pregnancy is confirmed, the couple are offered the option of terminating the pregnancy. Appropriate counselling should be provided at all stages of the screening process.

The aim of antenatal screening for infectious diseases, such as HIV, is first to detect the disease in the mother. If the disease is detected, the mother is informed of the likelihood that the child will become infected and interventions are offered that reduce the risk of transmission.

Differences in the screening tests available for detecting disease, the screening process and the long-term outcomes have implications for the modelling approaches for evaluating alternative screening options. The following sections review identified modelling studies in each of the three antenatal screening categories.

Autosomal recessive disorders

Autosomal recessive disorders require that both parents carry the defective gene (being either asymptomatic carriers of the gene, or being affected by some form of the disorder themselves) for the pregnancy to be at risk of the disorder. If the combination of carrier traits held by the parents is indicative of a serious disorder, then the pregnancy has a one in four chance of being affected by the disorder, and a one in two chance of being a carrier of a defective gene. The main disorders of this type are the haemoglobinopathies – thalassaemia and sickle cell disorder (SCD), and cystic fibrosis. The following section reviews modelling studies of antenatal screening for haemoglobinopathies, an area which is addressed in more detail in Chapter 8.

Review of haemoglobinopathies antenatal screening models

Haemoglobinopathies are serious conditions that result in reduced life expectancy and require regular treatment. The thalassaemic disorders primarily affect persons from an Asian or Mediterranean ethnic group, whereas the SCDs mainly affect persons from a black ethnic group. In the UK, where antenatal screening is conducted at a local level and the ethnic groups at risk generally comprise the minority of the antenatal population, the relevant policy question has been whether to restrict the offer of screening to those ethnic groups at highest risk of having an affected pregnancy, or to adopt a universal screening programme in which all pregnant women are offered screening.[197] The additional benefits of universal screening are mainly due to increased coverage of women in non-north European ethnic groups who are falsely classified as being north European in a targeted programme, rather than identifying additional cases in north European ethnic groups.

Four cost-effectiveness analyses of antenatal screening programmes for haemoglobinopathies were identified, although two of the studies were based on directly observed data and did not use a model. Cronin and colleagues[198] describe the

cost-effectiveness of haemoglobinopathy screening at a specific screening centre in net benefit terms – the savings from the predicted treatment costs for affected pregnancies that are terminated as a result of screening minus the costs of screening. Phelan[199] reports the observed costs and detection rates for two multiethnic London communities.

Zeuner and colleagues[197] undertook a review of both antenatal and neonatal screening for haemoglobinopathies, with the objective of informing an economic analysis of options for both forms of screening. The framework for the economic analysis was a model with the characteristics of a decision tree, but which was implemented in a general programming language (SAS) due to the level of model complexity. The model has since been replicated in an Excel spreadsheet as a more conventional decision tree model, although programming is still required to analyse the model for the full range of ethnic groups and antenatal populations.

The model splits the antenatal population into 12 separate ethnic groups, each of which has a probability of carrying one of six significant haemoglobinopathy traits, or of being a non-carrier. The model then allows for the probability that a woman may have a partner from any of the 12 ethnic groups, and that the partner also has a probability of carrying one of six significant haemoglobinopathy traits, or of being a non-carrier. The model, therefore, estimates the number of pregnancies in each of the 7056 (12 × 7 × 12 × 7) subgroups determined by possible combinations of parental ethnic status and haemoglobinopathy carrier state.

Outcomes included the number of live births, termination of pregnancies (TOPs), PND miscarriages and other pregnancy losses for both affected and normal pregnancies. The model was validated by comparing outputs from the model to previously predicted numbers of affected fetuses in the aggregate UK population (by type of haemoglobinopathy), and also comparing the number of PNDs predicted by the model with data recorded on the PND register (also by type of haemoglobinopathy). Baseline results were presented for three hypothetical antenatal populations comprising populations with a low, medium and high prevalence of mothers at high risk of carrying haemoglobinopathy traits. Data from the 1991 Census describing the proportion of the antenatal populations in each of the defined 12 ethnic groups were used to estimate the cost-effectiveness of universal versus selective screening for individual health authorities.

Gallivan and colleagues[200] present a feasibility study of a stochastic modelling framework for the evaluation of antenatal screening for haemoglobinopathies. Equations are presented to estimate the frequency of alternative fetal genotypes for different combinations of maternal and partner genotype, as illustrated by the following equation that describes the probability of an affected pregnancy when both the mother and her partner are carriers of the recessive gene:

$$\vartheta_{i,j} = \beta_i \varepsilon_j \left[\frac{1}{4}\Gamma_i + \sum_{k=1}^{R} \chi_{i,k} \left(\frac{\varepsilon_k}{4} + \frac{\varsigma_k}{2} \right) \right]$$

where $\vartheta_{i,j}$ is the probability of an affected pregnancy with a mother from risk group i and her partner from risk group j, β_i is the probability that the mother carries the recessive gene, ε_j is the probability that the partner carries the recessive gene, Γ_i is the probability that the partner is the true biological father, $\chi_{i,k}$ is the probability that the biological father of a mother from risk group i is from risk group k and is different from the declared father (the partner) and ε_k and ς_k are the probabilities that the biological father (if different from the declared father) is a carrier (heterozygous) or a sufferer (heterozygous) of the recessive condition, respectively.

Further equations are specified that integrate the predicted occurrence of alternative fetal genotypes with the provision of screening to estimate a range of performance measures (e.g. waiting times between screening stages, gestational ages at diagnosis) and outcomes (e.g. affected TOPs, miscarriages, false positives and negatives).

The description of the gene frequencies is not fully developed as the model assumes the presence of only a single recessive gene (which if present in both parents may lead to a serious disorder), rather than accounting for the wide range of haemoglobinopathy traits, and the combinations of the different traits (between parents) that may lead to serious disorders. The derived equations would need to be considerably more complex to handle the full range of gene frequencies.

The structure of the above mathematical model and the former decision analytic approach are similar, although Gallivan and colleagues explicitly incorporate time-dependent parameters; for example, uptake rates for PND and TOP are a function of the gestational age at the time of offer,

and the costs incurred by pregnancies that spontaneously miscarry are a function of the gestational age at the time of miscarriage. However, it may be possible to adapt the decision tree approach to incorporate these factors, and the alternative methods for implementing the same model structure for the evaluation of antenatal screening for haemoglobinopathies are compared in Chapter 8.

Chromosomal abnormalities

Conditions caused by an abnormal chromosome constitution involve extra or missing chromosome material (either a whole chromosome or a chromosome segment). Turner's syndrome, fragile X syndrome and neurofibromatosis type 1 are all forms of chromosomal abnormalities, although Down's syndrome is the most common form for which antenatal screening is undertaken. The method of screening for Down's syndrome involves estimating the level of risk through the collection of one or more types of information from the mother, and/or ultrasound examinations of the fetus. Relevant information collected from the mother includes her age and one to three biochemical tests of the maternal serum (single, double or triple test). The effectiveness of the biochemical serum tests is a function of the accuracy with which gestational age is known, so ultrasound measurement of gestational age may also be part of the screening process. Screening ultrasound for chromosomal abnormalities informs the nuchal translucency (NT) test, which measures the thickness of the translucent band at the nape of the fetus's neck. Validated equations are available to estimate the level of risk based on combinations of screening tests. Normally, a threshold level of risk is defined (for example, a one in 250 chance that the pregnancy is affected), and women whose pregnancies are estimated to be at a higher risk than the threshold are offered PND.

Review of Down's syndrome antenatal screening models

The earliest identified analysis of antenatal screening for Down's syndrome is a cost–benefit analysis that was published in 1987, which compared the costs of screening with the estimated "averted excess lifetime costs of Down's compared to non-Down's individuals".[201] The number of affected births prevented was based on the number observed in a regional screening programme in the UK that offered PND on the basis of age alone. Referenced estimates of the effect of combining serum α-fetoprotein

measurement (single test) with age to define risk levels were used to inform the comparator. The estimated lifetime costs for individuals with and without Down's syndrome included lost parental output, own output, consumption and healthcare use, education and fostering costs. Alternative rates of replacement of terminated Down's syndrome pregnancies with healthy children were tested in the sensitivity analysis.

Shackley[202] compares triple serum testing versus the offer of amniocentesis on the basis of age alone. Local data described age-specific birth rates and Down's incidence rates (and, hence, the effectiveness of an age-only screening test). Sensitivity and specificity rates for the triple test were based on a risk threshold of one in 250, and applied to the described incidence rates to estimate the number of detected Down's pregnancies and falsely identified normal pregnancies. The analysis includes the direct costs of providing screening and the averted excess lifetime costs of Down's syndrome compared with normal individuals. The latter includes lost parental productivity, excess consumption, educational needs, additional NHS costs, capital costs (e.g. home adaptations), adoption and fostering costs and lost individual productivity. The main result is the average cost per Down's birth avoided (including and excluding averted lifetime costs), although the number of amniocentesis-related miscarriages is also presented.

Ganiats and colleagues[203] compare the double and triple tests with age-based screening using a simple decision tree approach. The tree structure is not presented, but the process is described as all women are offered screening, all positive screens are offered PND (amniocentesis) and all positive PNDs are offered TOP. It is likely that the tree differentiated between women aged under 35 years and those aged 35 years and over, as the input data describes separate estimates of the number of births, the number of Down's cases and the sensitivity rate of the triple test for these two age groups. The results of the model are presented solely as the cost per affected case detected. The impact of screening on the miscarriage rate among unaffected pregnancies is raised in the discussion, which concludes that the double and triple tests reduce the number of amniocenteses undertaken, and hence the number of screen-related miscarriages (assuming 100% take-up of PND among women aged 35 years and over in an age-alone screening programme!).

Fletcher[204] also uses a decision tree to compare five alternative screening options based on the offer of biochemical testing with ultrasound measurement of gestational age, direct amniocentesis or both (with biochemical testing being paid for by the mother) for different age categories of the antenatal population. *Figure 15* describes the tree structure for women offered biochemical testing. The tree describes the offer and uptake of the biochemical test screening strategies, followed by the definition of high or low risk for those women undergoing screening. Uptake of amniocentesis is described for the high-risk group, which may lead to procedure-related miscarriage, or a positive or negative result. The tree stops at the detection of a Down's syndrome pregnancy, whereas the negative PND results may miscarry after the PND, or lead to a live birth of a Down's and 'other' child.

Local data sources informed age-specific pregnancy rates, Down's syndrome prevalence and screening acceptance rates. Sensitivity and specificity rates for the triple test at different risk thresholds were based on published estimates. Financial costs to the NHS were measured (tests, amniocentesis, initial and further counselling, TOP), although averted lifetime costs were excluded because the local health authority had indicated that it would not take downstream costs (i.e. of prevented affected births) into account.

Validation involved a comparison of estimated programme costs, the number of Down's births and TOPs, and the number of amniocenteses undertaken with observed numbers in the local area. The following outcomes were collected: live Down's birth, live other birth, TOP of Down's fetus, amniocentesis-related miscarriage, Down's miscarriage, other fetus miscarriage. Fletcher raises doubt over methods for measuring aggregate preferences, preferring the presentation of separate outcomes leading to debate amongst decision-makers, thus no cost-effectiveness ratios are estimated.

Wald and colleagues[205] use a decision tree-like approach to estimate the number of Down's births avoided and unaffected fetal losses. Unlike most decision tree applications, separate pathways are described for Down's syndrome and unaffected fetuses. The total number of Down's fetuses in the antenatal population was estimated, to which a detection rate (sensitivity) was applied. Acceptance rates for amniocentesis, and then TOP, and also the rate of spontaneous miscarriage for Down's syndrome pregnancies were specified. The model did not explicitly account for age effects; for example, an aggregate detection rate was assumed, although the assumption of different amniocentesis acceptance rates implicitly adjusts for a difference in the age distributions of the mothers of affected and unaffected fetuses. The results of the analysis are presented as both the ratio of Down's births avoided to fetal losses and the cost per Down's birth avoided. Extensive sensitivity analyses are presented for both results.

Vintzileos and colleagues[206] compared the early application of the NT test and subsequent offer of
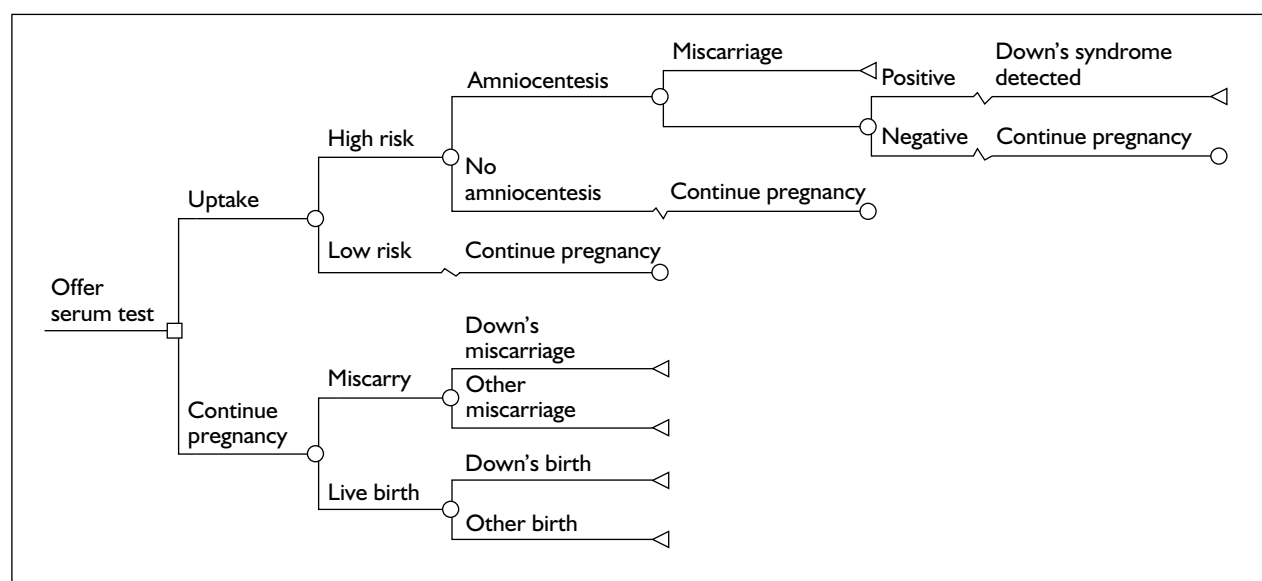

FIGURE 15 *Decision tree structure describing screening process for women offered biochemical testing (adapted from Fletcher[204])*

CVS to women at high risk, with a policy of offering CVS to all women aged 35 years and over. A simple equation was used to estimate age-specific thresholds for the required sensitivity of the NT test such that the net costs of both strategies were equal, given age-specific risk levels (e.g. one in 210 for women aged 35 years) and a range of false positive rates (5, 10 and 15%). The equation included the costs of the ultrasound required for the NT test, CVS PND and the lifetime cost of children born with Down's syndrome. The last cost included the costs of medical, developmental and special educational services and also the lifetime costs associated with lost productivity.[207] Baseline estimates for all parameters are also used to estimate the average cost per case detected, number of fetal losses (CVS-related miscarriages), and net benefits assuming 100% uptake of CVS and TOP.

Beazoglou and colleagues[208] estimated net benefits for six combinations of available biochemical tests (three single tests, two double tests and one triple test). The triple test was evaluated with and without ultrasound-based gestational dating. Net benefits were estimated using the following equation, which relaxes some of the assumptions implied by the same group:[206]

$$NBi = (I\ Si\ Up\ A\ Ab\ Su\ B) - (Pi + Pa\ Up$$
$$[Si\ I + Fi\ (1 - I)] + I\ Si\ Up\ A\ Ab\ Pga)$$

where I is the number of affected pregnancies in the 2nd trimester, Si is the sensitivity of screening test i, Up is the uptake rate for amniocentesis, A is the fetus survival rate following amniocentesis, Ab is the Down's syndrome termination rate (following a positive PND), Su is the Down's syndrome survival rate to term, Fi is the false-positive rate for screening test I, Pi is the cost of screening test i, Pa is the cost of amniocentesis and Pga is the cost of a Down's syndrome termination.

The prevalence of Down's syndrome in the second trimester was estimated using data describing the number of live Down's syndrome and unaffected births, and the estimated survival rate for Down's syndrome and unaffected pregnancies from the second trimester to term (it is not clear how the existing termination rate for Down's syndrome pregnancies was incorporated). Women aged under 35 years and aged 35 years and over were analysed separately. Age-specific sensitivity and specificity rates were used for the triple test (implying that age-specific rates were not available for the other test combinations). The results are presented as the net benefits of alternative

screening tests, although the number of amniocenteses undertaken is also presented. Sensitivity analyses describe the elasticity of individual parameters – the percentage change in the net benefit in response to a percent change in an input parameter value.

Benn and colleagues[209] used the above equation[208] to evaluate the use of ultrasound compared to the use of time since last menopause (LMP) to estimate gestational age, to improve the accuracy of the interpretation of the biochemical tests of maternal serum. Age-specific sensitivity and specificity rates for the ultrasound and LMP approaches were estimated by sampling random triple test results from trivariate Gaussian distributions of biochemical test results in normal and Down's syndrome pregnancies (informed by separate datasets for each approach). The relative probability that a triple test result was from an affected pregnancy (the likelihood ratio) is estimated. The risk threshold was set at one in 270, and the proportions of affected pregnancies above the threshold (sensitivity), and normal pregnancies below the threshold (specificity) were estimated for both dating methods.

Seror and Costet[210] compared the double and triple test for Down's syndrome. Gaussian distributions of the marker values for both tests for normal and Down's pregnancies were estimated using data describing the test results of over 10,000 non-Down's cases and 63 Down's cases. The proportion of each population that would be defined as a positive screen for different marker values informed the sensitivity and specificity of alternative PND policies (e.g. offering amniocentesis to the 5% of women at highest risk). Detection rates and false-positive serum screening results are also presented for a risk threshold of 1:250 based on observed numbers of live births for each maternal age.

Direct cost estimates were based on the resources involved in the production of 'double' and 'triple' tests by describing an optimised production process, assuming resources necessary to increase production were always available and that all equipment was dedicated to serum screening. Avoided lifetime treatment costs (based on a lifetime of care in a special hospital) were included in an analysis of 'societal profitability'. The risk thresholds for each test were varied to identify the maximum net benefits for each test. Further analysis incorporated a 'fetal loss risk aversion factor' – the value attributed to the loss of an unaffected fetus compared to the value attributed

to a diagnosed case of Down's syndrome, where the latter value is apparently defined as the avoided treatment costs. Seror and Costet state that a risk threshold of one in 250 implies that the decision-maker attributes the same value to the birth of an affected child as to the miscarriage of a healthy fetus. It is not clear if this statement is simply based on an assumed miscarriage rate of 0.4%, or whether a more complex calculation was devised.

Christiansen and Larsen[211] evaluated the double test followed by either immediate diagnostic test (CVS or amniocentesis) (e.g. if risk >1:65), no further test (e.g. if risk <1:1000), or further screen by NT measurement (e.g. if risk <1:65), compared with alternative screening strategies in the first and second trimesters. Separate distributions of serological markers and NT in the first and second trimesters are identified. Presumably, the NT distribution is correlated to the marker distributions as Monte Carlo simulations were undertaken in which values were sampled from the relevant marker distribution and the NT distribution, to allow the estimation of the final risk for a particular pregnancy. Risk values based on the sampled serological (double) test marker (*a*) were combined with the likelihood ratio (Down's syndrome:normal fetus) derived from the NT measurement (*r*) to give the final risk level (*ar*). Using the NT distribution and the equation for the distribution of the NT log(MoM) (multiples of the median of normal values – the statistic used to calculate risk levels) in normal and Down's syndrome pregnancies, a threshold value of *a* is established that any observed NT test result will not raise the final risk above the specified threshold (in this case, one in 400), hence NT measurement is not undertaken. Similarly, an upper risk threshold for *a* is estimated, above which any observed NT test result will not lower the final risk below the specified threshold. Direct screening and associated costs are presented, and also a conservative estimate of £35,000 for the treatment and care costs per case of live-born Down's syndrome, although the averted costs are not included in the presented ICERs.

Gallivan and Utley[212] present an illustrative mathematical model of the cost-effectiveness of NT screening, followed by CVS for PND, which included equations for estimating various screening-related outcomes. Six pregnancy categories were defined: normal or Down's live birth, normal or Down's spontaneous miscarriage and normal or Down's TOP not related to Down's. For each pregnancy category, equations were

defined to estimate the following outcomes, which are all described as a function of gestational age: NT tests correctly indicate the presence of Down's syndrome, NT tests wrongly indicate the presence of Down's syndrome, CVS indicates the presence/absence of Down's or causes miscarriage as a function of gestational age. The following equation describes the probability that CVS has been undertaken and indicated the presence of Down's syndrome by time *t* during gestation (a zero false-positive rate is assumed for CVS, so the number of detected cases of Down's is described):

$$W(t) = \sum_{i=1}^{6} \vartheta_i \alpha_i n_i(T_N) a_i(T_C) q_i(T_C)$$

where ϑ_i is the proportion of pregnancies in each of the six pregnancy states described above ($i = 1, \ldots, 6$), α_i is the screening uptake rate, $n_i(T_N)$ is the probability of a positive NT test at time T_N, $a_i(T_C)$ is the probability that a positive CVS is undertaken at time T_C and $q_i(T_C)$ is the probability that a fetus is alive at time T_C.

No results are presented, although it is stated that the approach is stochastic, using probability techniques to reflect patient-to-patient differences, diagnostic inaccuracies and the occurrence of adverse events. There is also the potential for applying analytical optimisation methods.

Infectious diseases

Antenatal HIV screening is the principal example of an antenatal screening programme for an infectious disease. The screening process differs from that described for the other forms of antenatal screening in that the objective of screening is not to identify an affected fetus, and possibly to terminate affected pregnancies. Antenatal HIV screening does not require prenatal diagnosis of the fetus; diagnosis is only required of the mother. Although it is possible that women who are identified as HIV positive may choose to terminate their pregnancy, the primary objective of antenatal HIV screening is to prevent the vertical transmission of HIV to the infant. Modelling the cost-effectiveness of antenatal HIV screening to the point of diagnosis is simpler than for autosomal recessive or chromosomal abnormalities, as only the mother is tested, although more detail may be required post-diagnosis in order to model the effectiveness of the preventive measures adopted. The following section reviews modelling studies of antenatal screening for HIV.

Review of HIV antenatal screening models

Ecker[213] evaluated antenatal HIV screening using decision analysis, although the tree structure is not presented. It appears that a simple structure describes the acceptance rate for screening, followed by the four screening outcome states (true and false positive and negative). False-negative and unscreened HIV-positive women have a probability of vertically transmitting HIV. The relative risk of transmission associated with zidovudine is applied to the probability of transmission for true positive women. The probabilities used in the model were derived from the literature, and ranges are presented for each parameter. Screening costs, treatment costs for false-positive women and the costs of early treatment for HIV are included (on the assumption that women may not have presented until later in the course of the disease without screening). The lifetime costs of treating an HIV-positive infant are also included.

Nakchbandi and colleagues[214] compared the effectiveness of mandatory with voluntary HIV testing of pregnant women. A decision tree describes the probability of opting out of PNC for both screening programmes (leading to alternative perinatal death rates), and women accepting or rejecting HIV testing under a voluntary programme. Input parameter values were obtained from the literature or estimated (by experts, presumably). Utility values were attached to each of the three outcomes: 1 for a healthy infant, 0 for a dead infant and 0.1 for an HIV-infected infant (estimated by dividing the median life expectancy of an HIV-affected newborn by average life expectancy). The results are presented as the threshold deterrence rate (the percentage of women deterred from seeking PNC due to mandatory HIV testing) at which voluntary screening increases utility. The number of infants spared HIV infection is also presented. This model did not include effects on the mother of knowledge of her HIV status, nor the impact of other transmission preventative actions.

Myers and colleagues[215] also compared voluntary with mandatory testing for HIV during pregnancy (both following universal counselling). However, this study assumes that mandatory testing has no effect on the uptake of PNC, and that the same proportion of women testing positive under both screening options would accept zidovudine treatment (though these assumptions are tested in the sensitivity analysis). Termination rates were also assumed to be unaffected by a positive result.

The tree is not presented, but the model appears to describe sequential test results for the enzyme-linked immunosorbent assay (ELISA) and the Western blot screening test. HIV-positive women have a probability of transmission, to which a relative risk reduction is applied for true-positive women. The model also differentiates between women presenting before and after 34 weeks (lower treatment effectiveness is assumed for later presenting women). A similar range of costs to those used by Ecker[213] are included (costs of early detection, false positives and lifetime treatment for HIV-positive infants).

Ades and colleagues[216,217] estimated the cost-effectiveness of universal versus selective (high-risk) antenatal screening for HIV from a UK perspective (the previous papers were all from a US perspective). The technical report describes a summary tree that describes the pathway of the aggregate antenatal population, including nodes describing whether HIV status is known, unknown HIV is detected, pregnancies to HIV-positive women are terminated and continued pregnancies result in an HIV-positive child. Two sub-trees describe the identification rate for unknown HIV-positive pregnant women and outcomes of interventions to reduce transmission, respectively. The identification sub-tree splits the population into four risk groups, each of which has a separate prevalence and screening compliance rate. The intervention sub-tree describes the proportions of women who accept each of eight combinations of three interventions (zidovudine therapy, Caesarian delivery and bottle feeding). The identification rates and transmission rates for each intervention combination (including no intervention) for the four groups were combined and fed back into the main model to estimate the aggregate outcomes for the full antenatal population.

Paediatric and maternal costs and life-years were combined to estimate the net benefit of antenatal HIV screening, assuming that the decision-maker is willing to pay £10,000 to gain an additional life-year. The impact of screening on the prognosis of detected and undetected HIV-positive mothers and also the lifetime effects of screen-detected and non-screen-detected HIV-infected infants are included. This study incorporates the costs associated with a higher termination rate in detected HIV-positive women, although the impact of the terminations on life-years lost and treatment costs avoided are not included.

Postma and colleagues[218] estimated the cost-effectiveness of universal voluntary HIV screening

in the UK. Outcomes included the life-years gained from earlier treatment of detected HIV-positive women (1 year, reflecting the benefit of starting antiretroviral therapy earlier), and in infants in whom HIV infection is avoided (70 years, assuming the life expectancy of an HIV-infected infant to be 7 years). Methods for estimating the effectiveness of screening in detecting HIV-positive women and in reducing transmission rates are not described, although a model to estimate the lifetime costs of treating HIV-positive infants is described, which includes four stages of HIV infection (indeterminate, asymptomatic, symptomatic and AIDS). The time spent in each state is assumed to be exponentially distributed, indicating the use of a Markov model. The benefits of following HIV-infected infants from birth are not included.

Zaric and colleagues[219] compared voluntary antenatal screening, neonatal screening and antenatal plus neonatal screening with current practice (where HIV screening can be requested), including the option of providing the newborn infant with antiretroviral therapy to reduce the probability of transmission. A decision tree described the pathway of pregnant mothers accounting for the proportion of HIV-infected women who have prior awareness of their status, the proportion accepting PNC, and then accepting HIV screening. The tree described six outcomes: mother HIV+ treated early, infant HIV+ treated early; mother HIV+ treated early, infant HIV+ treated late; mother HIV+ treated late, infant HIV+ treated late; mother HIV+ treated early, infant HIV–; mother HIV+ treated late, infant HIV–; and mother HIV–, infant HIV–. Life expectancies for infected and non-infected newborns and for infected mothers are referenced. The base case assumed no improvement in life expectancies for infected mothers or infected infants due to early detection, although improvements of up to 4 years were tested in the sensitivity analysis.

Immergluck and colleagues[220] evaluated (implied mandatory) universal, voluntary, and no antenatal screening for HIV in Chicago. The study described transmission rates in the complete absence of antiretroviral therapy, when both mother and infant receive therapy, and when only the infant receives the intervention. The impact of breast feeding is not included. A decision tree model described the pathway of patients accepting or rejecting screening, their subsequent screening outcome state (true or false positive or negative), the acceptance of treatment and the associated

probabilities of infection. Average life spans and costs for affected and unaffected infants were added to the respective end-points of the model.

Discussion

The review of antenatal screening models identified that the following outcomes were usually reported:

- birth of an unaffected child
- birth of an affected child
- termination of an unaffected pregnancy due to false-positive test results
- termination of an affected pregnancy
- PND-related miscarriage of an unaffected pregnancy
- PND-related miscarriage of an affected pregnancy.

Additional outcomes included total costs of the screening programme, the provision of informed choices to parents with affected pregnancies and the number of cases of unnecessary anxiety caused by false-positive test results.

The combination of the relatively short time horizon and a range of relevant outcomes has made the decision tree the most commonly applied modelling technique for the evaluation of antenatal screening. Some studies have not explicitly structured the antenatal screening process, using simple equations to estimate relevant outcomes, such as multiplying the disease incidence rate by the sensitivity of the screening test to estimate the number of affected pregnancies detected. These approaches involve the same calculations, although the decision tree estimates the range of outcomes simultaneously and provides an explicit representation of the process being modelled. Therefore, the decision tree may be preferred for simple model structures.

However, other studies have highlighted structural issues that may increase the complexity of the screening process beyond the feasible limits of a decision tree. Gallivan and colleagues describe possible time-dependent parameters in antenatal screening models,[200,212] primarily related to components of the screening process, such as screening test characteristics and uptake rates, which may vary as a function of the gestational age of the pregnancy. Zeuner and colleagues also found that a decision tree could not handle the necessary subgrouping of the eligible population.[197] In both these cases, more complex

mathematical models describing the relationship between input parameters and the relevant outcomes were adopted.

Earlier studies of Down's syndrome estimated the net benefits of screening by defining the benefits as the cost savings due to the avoidance of affected births,[202] though the processes for the estimation of lifetime costs were not explicitly presented. More recent modelling studies in these areas have tended to describe a range of short-term outcomes associated with antenatal screening,[204,205] such as affected pregnancies detected, and prenatal diagnosis-related miscarriages of affected and unaffected pregnancies. Where the informed offer of the termination of an affected pregnancy is a key objective, the estimation of conventional long-term outcomes such as (quality-adjusted) life-years gained may be considered inappropriate. There are complex moral and ethical issues around the effects of screening on the unborn child, the parents (and siblings, etc.) and subsequently born healthy children who may be regarded as 'replacements' for terminated affected pregnancies.[138] Currently, the preferred process of interpretation appears to be an implicit approach in which decision-makers are expected to apply their own weights to the range of screening outcomes. Modelling the lifetime treatment costs of affected children may inform decision-makers, though the accuracy of the prediction will diminish as the expected life expectancy increases.

The case for modelling the lifetime pathways of HIV-affected and unaffected infants is stronger for antenatal HIV screening than for the other forms of antenatal screening because the main recipient of the benefits associated with antenatal HIV screening is the infant who gains from the avoidance of the transmission of HIV. The gain to the parents from having a healthy child is analogous to the external effect on family and friends of improved health in any individual, which is not normally incorporated into utility-based measures of health outcome. The modelling of antenatal screening for HIV may be viewed as a hybrid of antenatal and neonatal screening – a model should describe the process of defining the risk of disease antenatally, whereas the effectiveness of interventions aimed at reducing the burden of disease should be modelled as a standard treatment model that predicts the long-term health benefits. As noted above, modelling lifetime treatment pathways for children is subject to vast uncertainties, and these uncertainties should be clearly expressed.

The benefit of some screening programmes for autosomal recessive disorders and chromosomal abnormalities may include the early detection of disease in affected pregnancies that are not terminated, for example, SCDs in a haemoglobinopathies screening programme. However, the description of the cost per QALY gained based on the proportion of pregnancies that are not terminated does not represent the full impact of the screening programme.

Chapter 8

Modelling antenatal screening programmes: a case study evaluation of antenatal screening for haemoglobinopathies

Introduction

The aim of the case study evaluation presented in this chapter is to compare the use of decision trees and analytical mathematical models for the evaluation of antenatal screening programmes. Decision trees are the simplest of the commonly used decision modelling techniques, and they are generally considered to be best suited to modelling situations in which the relevant events occur over a short period.[221] Antenatal screening primarily covers a period of less than 9 months, describing the maximum duration of a pregnancy. In the context of decision analytic modelling, this is a relatively short period, and the assumed use of decision trees to evaluate antenatal screening programmes appears reasonable.

Gallivan and colleagues[200,212] presented the use of analytic mathematical models as an alternative to decision trees for the evaluation of antenatal screening programmes. Mathematical models develop explicit equations to describe the relationship between input parameters and the outcome parameters of interest to a screening evaluation. Preliminary applications of the mathematical modelling technique have been developed for antenatal screening for Down's syndrome[212] and haemoglobinopathies.[200] These studies suggest that the analytic approach has several advantages, including the more manageable handling of complex antenatal screening programmes, and increased transparency as the influence of particular assumptions or parameter values is explicit in the specified equations. It is also stated that direct models may provide better insight into the structure of a problem and the relative importance of different factors.

Importantly, Gallivan and colleagues recognise that the alternative approaches are capable of incorporating exactly the same model structures, assumptions and data inputs. Thus, the relative advantages of the alternative modelling approaches will be process related, such as the ease with which the model is built and verified, and understood by users.

Criteria for comparison

The following criteria are defined, which are based on issues raised in various published guidelines for the conduct of decision analytic models:

- *Transparency*: It should be possible for a user to examine the structure of a model, so that they are able to understand the way in which the model works, and how the results are arrived at.
- *Interpretability*: The results of the model should be clear and interpretable for the decision that they are being used to inform.
- *Resource use to build the model*: Resource use comprises time and expertise. It is obviously a benefit if the same level of complexity can be represented using a simpler approach that requires less time to develop than the other.
- *Verification*: The process of verification ensures that the internal workings of a model are correct, that is, that the model estimates the expected outputs from whatever parameter values are input.
- *Time taken to run the model*: The time taken to analyse this particular model may be significant if probabilistic sensitivity analyses are required, or individual results for different localities are specified.

Background

The haemoglobinopathies are a range of genetically inherited disorders of red blood cell haemoglobin (Hb), which comprise the SCDs and the thalassaemias. The most common SCD is sickle cell anaemia, which is also usually the most severe form. SCDs can affect any organ in the body and produce a wide range of symptoms. The main symptoms of SCDs are anaemia, pain or infection. In the UK, the most common and important acute events include painful crisis (the episodic

exacerbation of pain, anaemia or jaundice), pneumococcal sepsis, splenic sequestration, acute chest syndrome, stroke and acute anaemia. In the worst cases, these symptoms can be life-threatening.

The most common type of thalassaemia is β-thalassaemia major, which usually presents in the first year of life with progressive haemolytic anaemia. α_0-Thalassaemia hydrops fetalis is relatively rare in the UK but it is associated with maternal morbidity and mortality during pregnancy and is almost invariably fatal *in utero* or shortly after birth.

Hb disorders follow an autosomal-recessive pattern of inheritance. This means that an abnormal Hb gene must be inherited from both parents for the child to be affected. There is a one in four chance of an individual being affected if both parents are carriers of an abnormal Hb gene. The six different abnormal Hb genes considered in this study are HbS, HbC, HbD, HbE, β-thal and α_0-thal. There are many other unusual Hb traits but these are very rare in the UK, and the six traits considered here represent the most commonly encountered types in the UK. If the fetus inherits a trait from only one parent then it is a carrier of the disorder (heterozygote) and will be healthy. Even if the fetus inherits a trait from both parents (homozygote or compound heterozygote), it might still be healthy, since many of the genetic combinations are not clinically significant.

The prevalence of all of the haemoglobinopathy traits varies significantly depending on the ethnic origin of an individual; these traits affect mainly black, Asian and Mediterranean ethnic groups. In general, all non-north European ethnic groups are considered to be at high risk. In a universal antenatal screening programme, all pregnant women are eligible for screening. A selective antenatal screening programme, however, aims to identify high-risk groups through a set of questions designed to establish the ethnic origin of the mother. Testing is offered to all women identified as being non-north European and/or with a known low mean corpuscular haemoglobin (MCH) level (which is measured automatically as a part of general antenatal obstetric care). A low MCH level (generally less than 27 pg) is associated with β-thal and α_0-thal traits. In effect, all women are screened for thalassaemia traits.[197]

Model structure

The same model structure is assumed for both modelling approaches. The following general assumptions have been made:

- The antenatal population is composed of 12 ethnic groups. They are based on the Census output classifications, which have been extended and are as follows: black Caribbean, black African, black Other, Indian, Pakistani, Bangladeshi, Chinese, Other Asian, Other, Cypriot, Italian and north European.
- All women and partners in the antenatal population are either non-carriers, or are carriers of one of six significant haemoglobinopathy traits: HbS, HbC, HbD, HbE, β-thal and α_0-thal.
- Heterozygous women and partners are not considered separately in the model because the additional complexity was considered disproportionate to the accuracy gained from modelling the small numbers of non-symptomatic individuals.
- The ethnic origin of the partners of women in the antenatal population is derived from a 12×12 matrix, that is, women from each ethnic origin may have a partner from any of the 12 ethnic origins.
- Each pregnant woman carries one fetus.

Figure 16 is a flow diagram showing the stages involved in the screening process by which pregnancies are defined. Women presenting after the end of the second trimester (26 weeks' gestation) are ineligible for screening, and non-north European women (with normal MCH) are ineligible for a selective programme. In addition to the misclassification of women in a selective programme, eligible women may not be offered screening due to a lack of knowledge and training in this area amongst the midwives or time restrictions.

Acceptance rates for screening vary by ethnic group for both mother and partner screening. Partner screening uptake rates are also allowed to vary by disease category, that is, SCD and thalassaemia. The screening test is highly accurate, although the model allows for the possibility of a false-negative screen result. If the father is not screened for any reason, then the mother passes straight to stage 9, and is offered a PND. The model allows for the possibility of the declared partner not being the biological father. If the declared partner is a carrier, then the couple proceed through the screening process and an affected fetus would not be missed. If the declared partner is either a non-carrier or a carrier of a trait that produces only non-significant combinations when combined with the mother's trait, then an at-risk pregnancy will be missed.
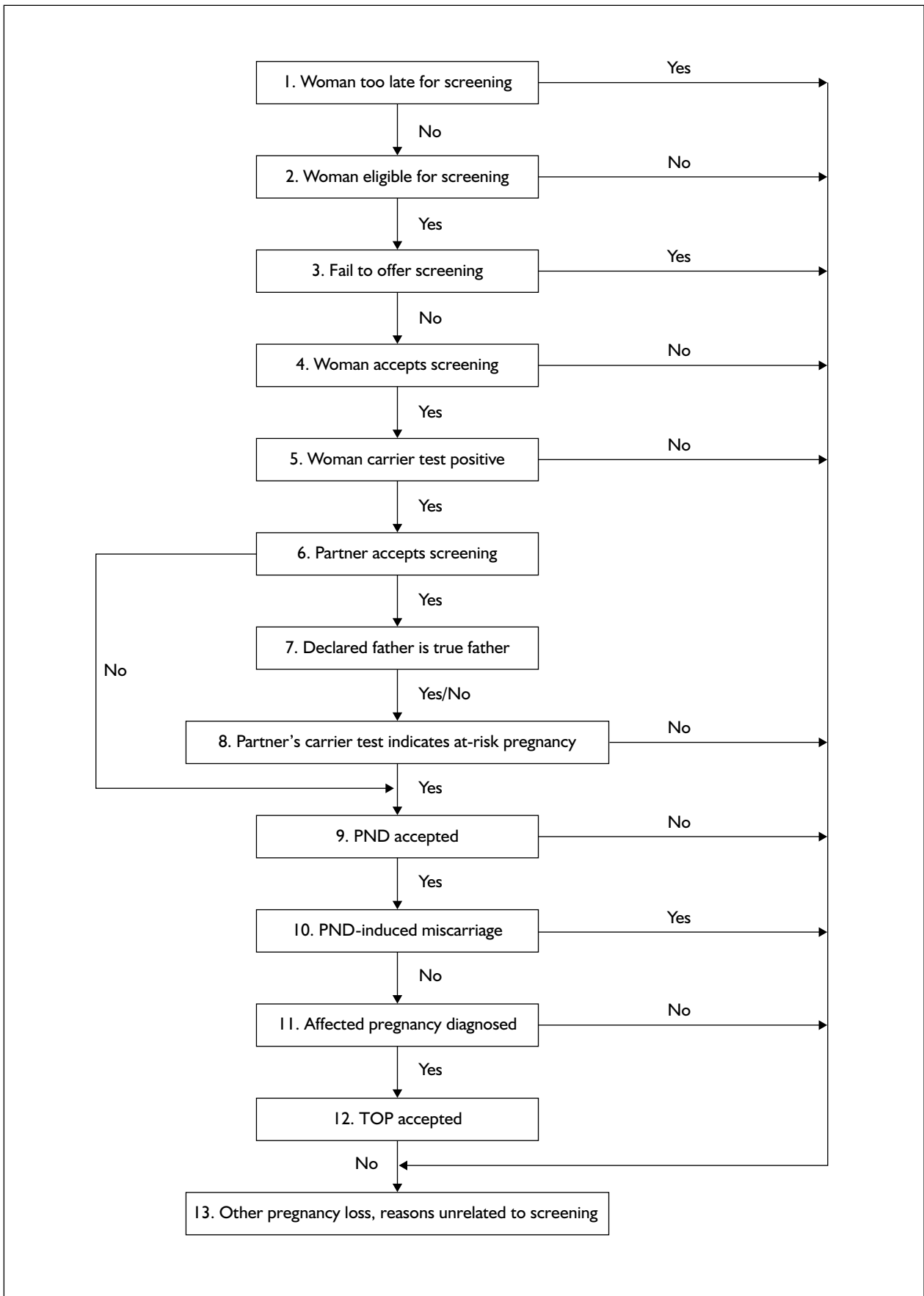
FIGURE 16 *Screening pathway for antenatal screening for haemoglobinopathies*

There are some data suggesting that the earlier the PND is offered during the pregnancy, the more likely a couple are to accept it. PND and TOP acceptance rates are described as a function of ethnic group, diagnosed disease and the time of offer. This means that the model can be used to investigate the effect of changes on the way in which screening is offered, such as it being offered in primary rather than secondary care.

The possibility of pregnancy loss for reasons unrelated to screening is described as the final point in the diagram. The model describes the probability that miscarriages occur at some point prior to each key stage (i.e. the maternal carrier test, the paternal carrier test, PND and TOP), such that a diminishing proportion of these pregnancies are assumed to survive as far as each of the four key stages, which affects the estimated screening costs.

Development of the decision tree

The following sections describe the different components of development process for the decision tree that was built to evaluate antenatal screening for haemoglobinopathies.

The decision tree pathways

The starting point for the decision tree is the actual risk status of the fetus. These initial stages are shown in *Figure 17*. The father referred to is the true biological father, so the risk status of the pregnancy is known at the start of the tree, but the possibility of the non-biological father being screened is considered at subsequent branches.

Three different 'sub-trees' follow. In the case where the mother is not a carrier, and the pregnancy is therefore not at-risk irrespective of the father's genotype, the remainder of the tree describes the probability that women are screened and that the non-affected birth continues to term (no false-positive test results are assumed), as shown in *Figure 18*.

For carrier women there are two separate 'sub-trees' depending on whether the pregnancy is at-risk or not. *Figure 19* presents an example of the sub-tree for pregnancies that are not at-risk as the true father is not a carrier of a significant Hb trait. The tree describes pathways leading to the termination of an unaffected pregnancy in cases where the non-biological father is screened and is found to have a significant Hb. *Figure 19* is

repeated six times, once for each of the significant haemoglobin traits. *Figure 20* shows the case where the pregnancy is at-risk, which is repeated seven times, to reflect all possible clinically relevant combinations of Hb traits.

Implementation

Microsoft Excel was chosen as the most suitable software in which to implement the decision tree, due to the unusually large size of the decision tree and the increased flexibility that Excel offers, compared with specialist software. The number of variables that affect the performance of the screening programmes resulted in many different branches of the tree and the large size of the decision tree was one of the main difficulties of implementation. If the decision tree was built conventionally as one big tree, then it would have been unmanageable. The tree needed to consider 12 ethnic groups for the mother and the father, plus seven possible haemoglobinopathy traits for each individual. Combining these parameters to estimate the true prevalence of affected pregnancies in an antenatal population would require 7056 branches. To overcome these problems the decision tree was adapted, and a solution was found using macros written in Visual Basic for Applications. A decision tree was developed that described the screening pathways for one of the 144 ethnic group combinations. For the first maternal ethnic group, the macro loops through each of the 12 possible paternal ethnic groups, collecting the results for each group as it goes along. This cycle of the twelve paternal ethnic groups is then repeated for the remaining 11 maternal ethnic groups.

The decision tree is contained in a single worksheet, which is linked to six other worksheets within the same Excel file that contain the input data for the model. At the end of each pathway in the decision tree, the overall path probability is calculated by multiplying the probabilities of every branch along that path. Adjacent columns calculate the path cost, by summing the individual costs that are accumulated along the path, and model outcomes, by summing the relevant path probabilities.

Verification

Verification describes the process of checking that the model produces outputs that are consistent with the parameters values input to the model. First, manual checks of the Microsoft Excel equations were undertaken to verify that each path summed to the correct outcome. This is a time-consuming and difficult task and the fact that the
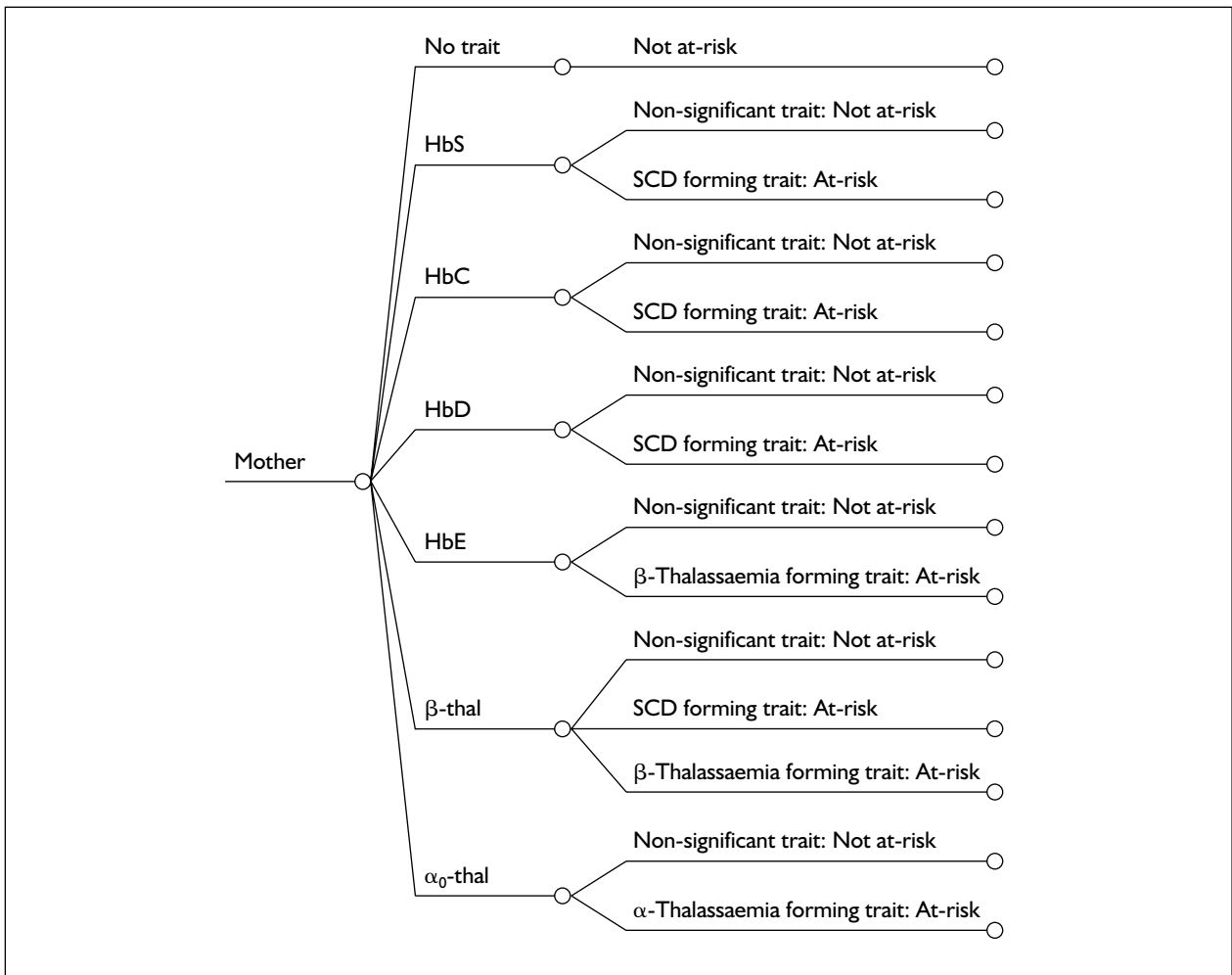
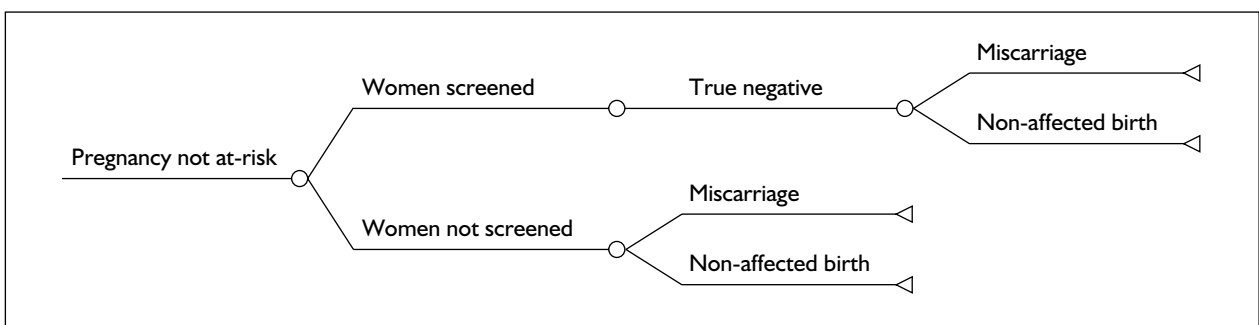**FIGURE 17** *Sub-tree: definition of mothers' carrier status*

FIGURE 18 *Sub-tree: mother not a carrier of haemoglobinopathy trait*

decision tree considers only one ethnic group combination at a time effectively means that the number of paths that require checking has been divided by 144, which is of great benefit.

Further verification involved setting all of the parameters to simplified values, the outputs for which could be estimated by hand, for example, the following input parameter set was estimated:

- Proportions of α_0- and α_+-thalassaemia carriers in the Chinese ethnic group set to one.
- Size of Chinese antenatal population set to 1000.
- All Chinese women have Chinese partners.
- All other ethnic groups contained no haemoglobinopathy traits.
- General failure to screen rates set to 0.5.
- Misclassification rates set to zero.
- Probability of miscarriage set to zero.

99
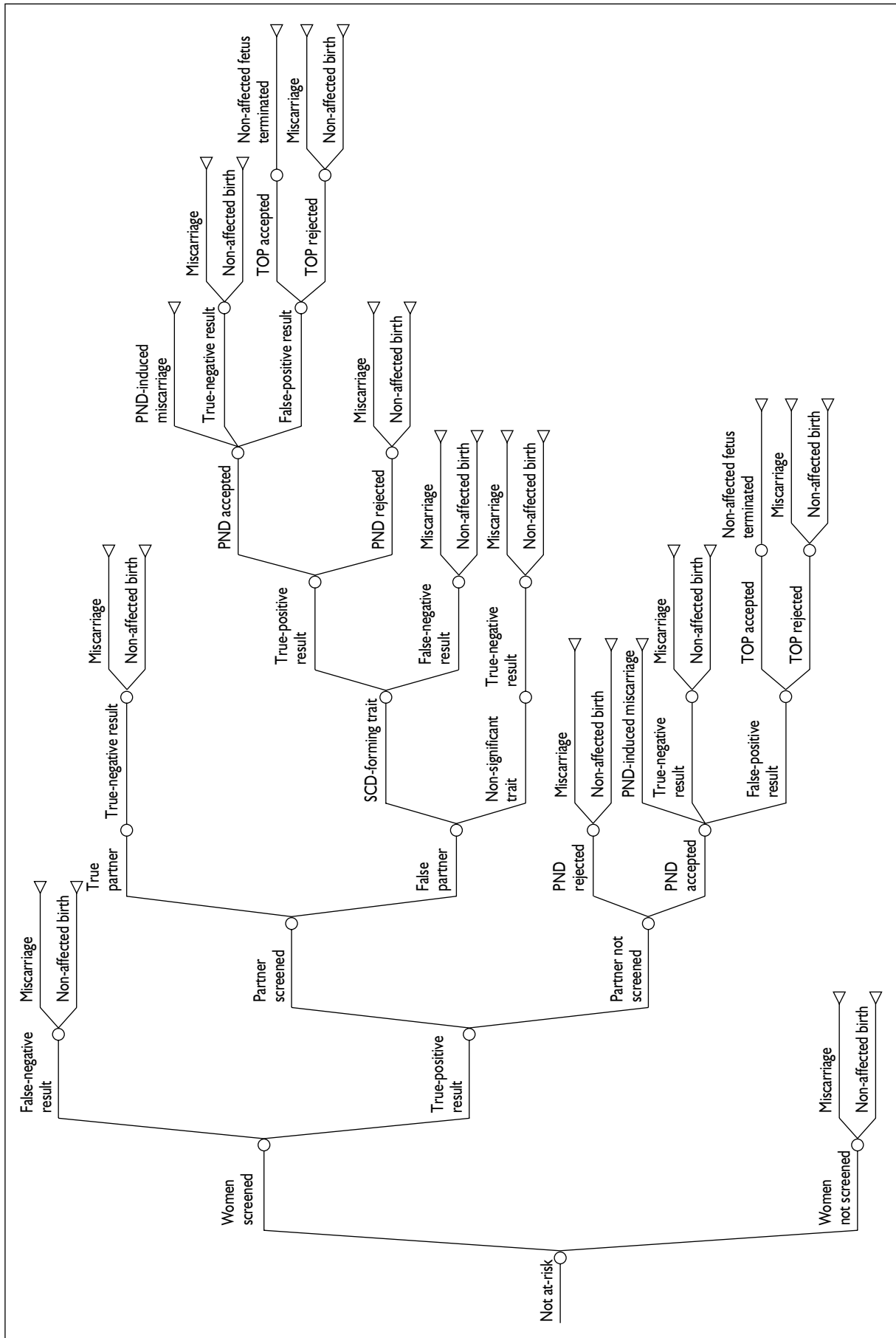
FIGURE 19 *Sub-tree: mother carrier, but pregnancy not at risk*
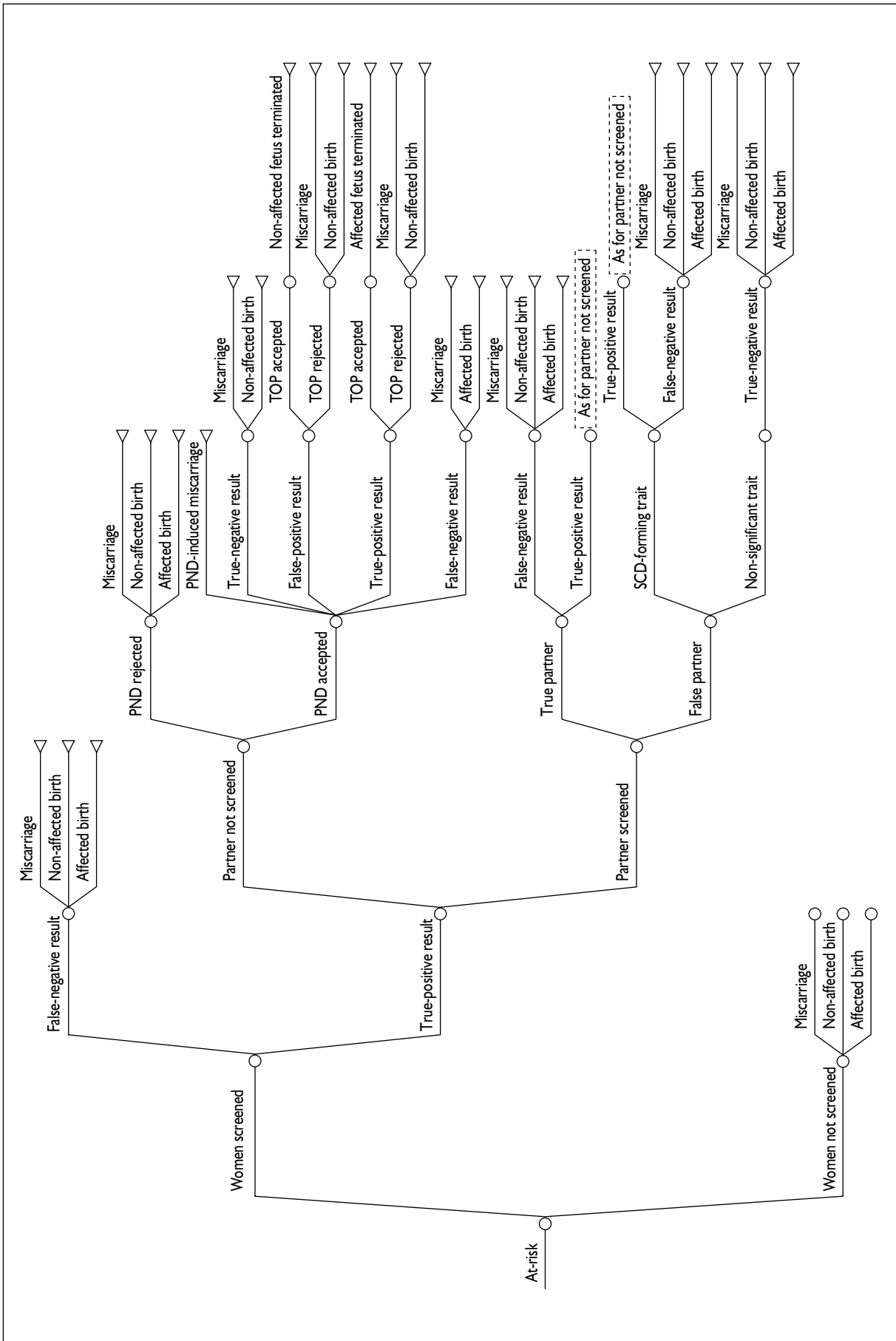
FIGURE 20 *Sub-tree: at-risk pregnancy*

Similar checks were also carried out by setting the proportions of the other traits to one in other ethnic groups, so that the α-thalassaemia, β-thalassaemia and SCD calculations could all be checked.

Development of the analytic model

Defining the model

The development of the analytic model was based on a theoretical model published as a working paper,[200] and also through discussions with the authors (Gallivan S, Clinical Operational Research Unit, University College London: personal communication, 2004). The CORU work on screening for haemoglobinopathies considered only one haemoglobinopathy trait, although it did account for heterozygous individuals with two haemoglobinopathy traits. The case study evaluation developed the pregnancy categories and equations presented in the CORU working paper to incorporate six different haemoglobinopathy traits, which even after excluding heterozygous individuals resulted in 92 pregnancy categories. An example of the equations for the probabilities of a fetus belonging to each of the 92 pregnancy categories is shown for the probability of a fetus belonging to category 2 (unaffected fetus, where the mother is a carrier of the HbS trait and the father has two normal Hb genes), when the mother is from ethnic group i and the declared father is from ethnic group j:

$$\theta_{2,i,j} = \\ Sf_i Am_j \left[1 - \sum_{k=1}^{R} \chi_{i,k} \left(\frac{Sm_k}{4} + \frac{Cm_k}{4} + \frac{Dm_k}{4} + \frac{\beta m_k}{4} \right) \right]$$

where Sf_i is the probability of a woman from ethnic group i being a carrier of HbS, Am_j is the probability of a man from ethnic group j not being a carrier, $\chi_{i,k}$ is the probability that the true biological father of a child of a mother from ethnic group i is different from the declared father and from ethnic group k and Sm_k, Cm_k, Dm_k and βm_k are the probabilities of a man from ethnic group k being a carrier of HbS, HbC, HbD and β-thal, respectively.

Pregnancies are divided into two categories, those which in the absence of screening would result in a live birth, and those that do not result in a live birth due to spontaneous miscarriage or stillbirth. Separate sets of analytical expressions were

developed to describe each of the following outcomes for births that proceed to term and for pregnancies that miscarry:

- maternal blood samples tested
- maternal blood samples tested with a positive result
- paternal blood samples tested
- paternal blood samples tested with a positive result
- PNDs
- PNDs with a positive result (total number and also split by disorder)
- positive PNDs followed by TOP
- PND-induced miscarriages
- affected fetuses lost due to a PND-induced miscarriage
- unaffected fetuses lost due to a PND-induced miscarriage.

Implementation

The analytic model was implemented in Visual Basic for Applications together with Microsoft Excel. The data are entered into the Excel worksheets and the calculations are completed by Visual Basic code, which outputs the results to separate Excel worksheets. There are several different subroutines within the Visual Basic code. The first few subroutines assign values to the parameter variables in the code by reading the input data from the worksheets. Another subroutine calculates the probability of a fetus belonging to each pregnancy category. *Figure 21* gives an example of how the pregnancy category equations were implemented in Visual Basic code.

Other subroutines calculate the frequency of the relevant outcomes, using the pregnancy category probabilities estimated in the previous subroutine. Separate subroutines were specified for universal screening and selective screening. Some of the initial calculations for universal screening were relatively simple, but they quickly become more complicated. The code for the selective screening calculations is more complicated again, as separate calculations were required for the north European ethnic group.

Verification

Two approaches to the verification of the analytic model were used. First, the code was written such that some of the outcomes were calculated in two different ways. For instance, the total number of true-positive PNDs was estimated as the sum of the true-positive SCD, β-thal and α-thal results, in addition to estimating the total number directly. Comparing the two results makes it possible to

```
' this calculates the probability of being in pregcategories 2–15
For i = 1 To 12
   For j = 1 To 12
      RunningTotal = 0
      For k = 1 To 12
      RunningTotal = RunningTotal + (NotRealFather (i, k) * (0.25 * (SmProb (k) + CmProb (k) _
      + DmProb (k) + BetamProb (k))))
      Next k

   PregCategory2 (i, j) = (SfProb (i) * AmProb (j)) * (1 – RunningTotal)
   PregCategory3 (i, j) = (SfProb (i) * AmProb (j)) * RunningTotal

   PregCategory4 (i, j) = (SfProb (i) * SmProb (j)) * (1 – (0.25 * IsRealFather (i)) – RunningTotal)
   PregCategory5 (i, j) = (SfProb (i) * SmProb (j)) * ((0.25 * IsRealFather (i)) + RunningTotal)

   PregCategory6 (i, j) = (SfProb (i) * CmProb (j)) * (1 – (0.25 * IsRealFather (i)) – RunningTotal)
   PregCategory7 (i, j) = (SfProb (i) * CmProb (j)) * ((0.25 * IsRealFather (i)) + RunningTotal)

   PregCategory8 (i, j) = (SfProb (i) * DmProb (j)) * (1 – (0.25 * IsRealFather (i)) – RunningTotal)
   PregCategory9 (i, j) = (SfProb (i) * DmProb (j)) * ((0.25 * IsRealFather (i)) + RunningTotal)

   PregCategory10 (i, j) = (SfProb (i) * EmProb (j)) * (1 – RunningTotal)
   PregCategory11 (i, j) = (SfProb (i) * EmProb (j)) * RunningTotal

   PregCategory12 (i, j) = (SfProb (i) * BetamProb (j)) * (1 – (0.25 * IsRealFather (i)) – RunningTotal)
   PregCategory13 (i, j) = (SfProb (i) * BetamProb (j)) * ((0.25 * IsRealFather (i)) + RunningTotal)

   PregCategory14 (i, j) = (SfProb (i) * AlphamProb (j)) * (1 – RunningTotal)
   PregCategory15 (i, j) = (SfProb (i) * AlphamProb (j)) * RunningTotal
   Next j
Next i
```

FIGURE 21 *Example of Visual Basic code used in the mathematical model*

perform a simple check on the calculations, and this process was repeated wherever possible throughout the development of the model.

Post-development verification involved the testing of convenient sets of input parameter values, for which the outputs of the model could be compared with output calculations by hand, as described in the verification process for decision trees.

Incorporating the time dependences

In the original decision tree, there was no allowance for the fact that PND and TOP acceptance rates may be dependant on the gestational age at which they are offered. It was, however, similarly straightforward to incorporate the additional complexities of the time-dependent uptake rates in the decision tree as in the analytic model. An extra worksheet was added to the Excel files of both models, which contained the probability distributions of the gestational ages at which different aspects of the screening process

are offered for each ethnic group. Estimates of the gestational age-specific uptake rates were combined with probability distributions describing the time-of-offer for PND and TOP to estimate a weighted uptake rate for PNDs and TOPs for each ethnic group. The estimated aggregate uptake rate was specified as a single input parameter.

The original model also assumed that miscarriages occurred at the end of the screening process, that is, all miscarriage pregnancies incurred all relevant screening costs. The analytic model links the probability distribution describing the time of miscarriage to the probability distributions describing the time of offer and associated uptake rates. A similarly external approach to the estimation of costs associated with miscarriages is applied to the decision tree model, whereby the full model is analysed for the pregnancies that continue to term and the costs incurred by the set of miscarriages were added to estimate the total costs. This approach actually simplifies the model structure as the 'spontaneous miscarriage' branches of the tree are removed.

Discussion

The main hypothesised advantage of the analytic alternative to decision trees for the evaluation of antenatal screening programmes was the improved handling of complicating factors, such as large numbers of branches, and the incorporation of time dependent parameters, such as miscarriage rates and acceptance rates for PND and TOP. To implement these factors, pragmatic adaptations to the decision tree approach were required, including the use of programming code to restrict the physical representation of the decision tree to one of a possible 144 sub-trees, and the external analysis of time-dependent parameters to inform aggregate parameter values. However, the decision tree retained the principle advantages of the approach, such as the high level of transparency and the explicit representation of possible pathways through the screening process. In contrast, for an unfamiliar user, the analytic model appears to produce the results almost from nowhere, since the model is completely contained in the Visual Basic code.

For this reason, if non-analysts want to be able to follow the logic of the model, then the decision tree approach may be preferred (although a basic understanding of Microsoft Excel would still be required). It may be beneficial to study user perspectives of both models, which could investigate the users' views with respect to self-analysis of the models, in addition to the understanding of written and verbal presentations of the models, particularly in a decision-making context. Alternatively, these issues may be context specific and should be considered at the point of contract between the research funder and the analyst.

There does not appear to be any significant difference in the time taken to build the two different models, although it is difficult to make a direct comparison since they were built separately by different analysts. The analytic model does have an advantage at the verification stage, however, since it is possible to calculate the outcomes in different ways in order to check that the results 'add up' (see the section 'Verification', p. 102). This means that many errors became immediately apparent during the development of the analytic model, but this was not possible in the decision tree. Although it may be easier to detect errors in the code of the analytic model, locating and correcting the errors is more difficult. It can often be due to something as simple as a bracket being in the wrong place, which can be very hard

to find in thousands of lines of code. It is often a simpler process in the decision tree, where the model consists only of the Excel equations in the cells. Therefore, although it is generally harder to detect errors in the decision tree, it is usually easier to correct them once they have been identified.

Another advantage of the analytic model is that it takes less time to run. This becomes particularly apparent when running the model for all local authorities in England and Wales. The analytic model takes approximately 15 minutes to do this, whereas the decision tree takes approximately 30 minutes. If the model is to be converted to a probabilistic model in the future, then an analytic model might be the best option. The impact of the running time advantage of the analytic model may also be related to use of the model; for example, a shorter running time may be a significant advantage for the interactive presentation of the model where alternative input parameter values suggested by the audience are tested on the spot.

On balance, it seems that neither approach is significantly superior to the other. The main advantage of the decision tree is that it has greater transparency than the analytic model, whereas the main advantage of the analytic model is that it has a shorter running time. It would be beneficial to carry out further research into user perspectives of the different models in order to identify how important the issue of model transparency is, and whether this is an over-riding concern. The significance of this factor will likely be affected by the extent to which the model is to be distributed, and analysed, by non-expert analysts.

None of the antenatal screening programmes reviewed in the previous chapter involved the same level of pathway complexity as the haemoglobinopathy models presented in this chapter. The evaluation of the full range of screening options for Down's syndrome involves different forms of complexity with respect to the time at which women are screened, due to the relationship between the combined sensitivity and specificity of different screening tests and the gestational age at which the tests are undertaken. The representation of these relationships in a decision tree would require the linked estimation of the test characteristics, and also the uptake rates for PND and TOP, using a similarly pragmatic approach as described for the haemoglobinopathies case study.

In general, it seems reasonable that the choice between the decision tree and the mathematical modelling approaches should be based largely on the personal experience and preferences of the analyst, as knowledge of model building and verification techniques may be viewed as the most important aspects of model development. Less complex model structures are easier to verify and will have insignificant running times, such that a decision tree may be regarded as the most appropriate modelling technique.

Chapter 9
Valuing screening specific outcomes

Introduction

This chapter reviews the published literature on valuing the benefits of screening. The benefits of screening programmes are frequently described in terms of the results of cost-effectiveness analyses, which have focused on benefits such as life-years gained.[222] Screening may have additional value by providing information.[223] This may be in the form of reassuring information of true negative status, or even if the information is not most favourable. Lange and colleagues[224] found that some women were willing to pay for information on autosomal dominant polycystic kidney disease (ADPKD) carrier status of their unborn child even if they were unwilling to abort the baby whatever the outcome. There may be less direct benefits from screening, such as feelings of empowerment and control due to participation.

Potential disbenefits to screening include temporary anxiety associated with the invitation and waiting for the results, or anxiety due to receiving a false-positive result. Mooney and Lange[225] suggest that among other possible disbenefits of screening programmes is that they force the population to make decisions they would not have had to make in the absence of the existence of the screening programme. For example, if a woman makes a decision not to make use of a screening programme for cervical cancer, and she is then diagnosed with cervical cancer, she may suffer regret from her previous decision. Mooney and Lange also point out that, when screening programmes are only available to certain populations, those who are excluded may feel deprived because they are aware that there is a screening programme that is not available to them.

There is also the possible anxiety that may be associated with the earlier diagnosis of disease. In extreme cases, individuals may never suffer symptoms because they die from another cause before symptoms occur. Harvei[226] commented that latent carcinomas of the prostate were found in 30–35% of autopsied men above the age of 50 years in Norway, and Etzioni and colleagues[227] estimated that around 75% of prostate cancers would never be diagnosed before death (although

this would evidently change if they were detected by screening). Most of the reviewed screening studies applied non-screening specific utility values to clinical health states, but one issue that needs to be addressed is whether screen detection alters the utility values associated with disease states.

The purpose of the review reported in this chapter was to determine the extent of previous research into the quality of life effects of the effects of screening in terms of true-positive, false-positive, true-negative and false-negative screening outcomes. Of particular interest were studies that obtained actual utility values or those that used generic health status measures that could be converted into utility values. The following section describes the search methodology, followed by the results of the search strategy. Included studies are then reviewed in detail to identify utility estimates that might be useful for economic evaluations in this area. The findings are discussed and recommendations made regarding the strengths and weaknesses of existing evidence and where future research might best be focused.

Methods

The search terms are described in Appendix 1, and included search terms that covered health state utility elicitation [including standard gamble (SG), time trade-off (TTO) and visual analogue scale (VAS)], preference-based measures of health (e.g. EQ-5D, HUI) and the Short Form with 36 Items (SF-36) (for which there are algorithms for deriving preference-based measures). The terms quality-adjusted life-years (QALYs) and disability-adjusted life-years (DALYs) were also used. The search was run in the Ovid MEDLINE database from 1966 to February Week 4 2004. The search detected a total of 350 papers. The abstracts were read and any papers deemed possibly relevant were selected and read. A total of 34 papers were selected in this way. The Harvard Catalogue of Preference Scores[228] was also searched, which identified a further 26 papers.

The papers selected from the above sources were obtained and read. Studies were included if they either attempted to measure utility values for

outcomes of screening or they provided values from a generic health status measure which could be converted to utilities.

Results

Out of a total of 60 papers, 10 comprised non-empirical discussion of issues relating to screening.[223,225,229–236] One study was a cost-effectiveness analysis exploring cost per life-year gained.[237] Six studies turned out not to be about screening.[238–243] A further 25 studies related to screening, but incorporated utility values either from the literature or author- or expert-based estimates.[165,166,168,174,244–264] A further study was unclear about the utility values they obtained,[158] and another study obtained non-screening specific disease state utilities.[265]

Seventeen studies involved the authors actually obtaining utility or health status values relating to screening outcomes, and were therefore considered relevant to this review.[125,222,234,266–279] The papers are summarised in *Table 33* in terms of screening disease, outcome, descriptive system, valuation method, valuation source and perspective of source (i.e. as users or hypothetically as prospective patients or members of the general population).

Descriptions of the relevant studies
De Haes and colleagues[267]
de Haes and colleagues used a sample of "experts" comprising 12 breast cancer experts and 15 public health employees to value 15 health states in the context of cases identified by breast cancer screening, including separate states for 'screening attendance' and the 'diagnostic phase'. The health state descriptions covered physical, psychological and social aspects and were based on a review of the empirical literature into the effects of screening. The states were valued by a VAS, in which 0 corresponded to the worst imaginable quality of life and 100 to the best imaginable quality of life. The authors converted VAS values to TTO scores using the function $1 - (1 - VAS) \times 1.82$. The median utilities were used for the analysis.

This was an interesting attempt to take account of the psychological consequences of screening for breast cancer, although the empirical link to data is questionable in places. The use of percentages in the descriptions is difficult to justify since these domains are not measured on scales with ratio properties. Another concern is that a sample of experts was used to value the states, which may

not be representative of the general population. The transformation of VAS into TTO is far from ideal and has been criticised in the economics literature (for a review, see Brazier and colleagues[281]). Nonetheless, the paper provides a direct attempt to quantify the importance of psychological outcomes alongside the more conventional clinical outcomes.

De Koning[266]
De Koning and colleagues refer to the same data as reported by de Haes and colleagues,[267] but they focused on true-positive and false-positive screening results, and therefore focused on treatment phases incurred as a result of screen-detected cancers, and further investigations of false-positive results. The same limitations arise as described for Haes and colleagues.

Dominitz and Provenzale[125]
Dominitz and Provenzale explored preferences for health-related quality of life (HRQoL) associated with screening for CRC. They interviewed four groups of patients who had not undergone CRC screening (62), about to be screened by FSIG (24), about to be screened by colonoscopy (114) and with CRC (46). Values were elicited using TTO to trade time in the described state for time in perfect health in units of days over a hypothetical life expectancy of 20 years for the following scenarios:

- 5-yearly screening for CRC by FSIG
- 5-yearly screening for CRC by colonoscopy
- 5-yearly screening for CRC by colonoscopy, but with one episode being complicated by perforation requiring surgery.

Units of years were traded for the following health states:

- living in the respondent's current health state
- living with colon cancer
- living with a colostomy.

The unscreened group were willing to trade more days to avoid all screening methods than the other three groups. The only significant difference in health state valuations between the four patient groups was between median values for current health given by the groups about to undergo colonoscopy screening and the colorectal cancer group ($p < 0.05$). The authors identified the high level of variability in valuations within each patient group, and suggested that preferences for screening could be very individual, and therefore different types of screening might suit different people, depending on their preferences.

TABLE 33 Utility studies included in the quality of life review

Study	Year	Screening topic	Outcomes	Descriptive system	Valuation method	Valuation source (*n*)	Perspective
de Haes et al.[267]	1991	Breast cancer	TP, FP	Health states	VAS transformed to TTO	27 experts	Experts
de Koning et al.[266]	1991	Breast cancer	TP, FP	Health states	VAS transformed to TTO	27 experts	Experts
Dominitz and Provenzale[125]	1997	CRC	Screening process	Vignettes	TTO	62 non-screened / 24 screened (FSIG) / 114 screened (colonoscopy) / 46 CRC	Patients / Non-patients / Screening participants
Cantor et al.[274]	1995	Prostate cancer	TP, FP	Health states	TTO	10 healthy men	General population
Volk et al.[275]	1997	Prostate cancer	TP, FP	Health states	TTO	10 healthy husbands / 10 healthy wives	General population
Krahn et al.[268]	1994	Prostate cancer	TP	Health states	TTO	10 physicians	Experts
Johnston et al.[222]	1998	Breast cancer	TP, FP, TN, FN	Vignettes	TTO and VAS	440 women of screening age	Users
Gerard et al.[279]	1999	Breast cancer	TP, FP, TN, FN	Vignettes mapped to EQ-5D	TTO	440 women of screening age	General population
Essink-Bot et al.[270]	1998	Prostate cancer	TN, FP	SF-36 and EQ-5D	Non-preference-based and TTO	600 screened / 235 non-screened	Patients / General population
Cormier et al.[276]	2002	Prostate cancer	TN	SF-36	Non-preference-based	220 relatives of cancer patients	Users
Edelman et al.[277]	2002	Diabetes	TN, TP	SF-36	Non-preference-based	1253 screened	Users
Hensley et al.[271]	2003	Ovarian cancer		SF-36	Non-preference-based	147 high-risk women	Users
Rumbold and Crowther[273]	2002	Gestational diabetes mellitus	TP, TN, FP	SF-36	Non-preference-based	158 pregnant women	Users
Fry et al.[272]	2001	Ovarian cancer		SF-36	Non-preference-based	27 oophorectomised women / 28 screened women	Users
Shackley and Cairns[269]	1996	Cystic fibrosis	*Ex ante* probabilities	Vignettes	SG	52 women	General population
Cairns et al.[278]	1996	Cystic fibrosis	*Ex ante* probabilities	Vignettes	SG	52 women	General population
Hall et al.[234]	1989	Breast cancer	TP	Vignettes	Healthy year equivalents (TTO)	44 women (non-patients) / 60 women (patients)	General population / Patients

FN, false negative; FP, false positive; TN, true negative; TP true positive.

This seems to be a well-conducted study using a recognised choice-based valuation technique. The only possible limitation is that the sample for some groups is rather small. It does not provide general population valuation, although these are likely to be closer to those of the unscreened group.

Cantor and colleagues[274]

Cantor and colleagues explored the effects of the treatment-related adverse events, incurred as a result of prostate cancer screening, including impotence, incontinence, gynecomastia (development of breasts), urethral stricture, rectal injury and death. The authors collected utility values for these states from a small sample of 10 married men of approximately 50 years of age and in good health, from a Texas medical centre. The TTO method was used to elicit these values, which identified impotence as having the highest mean value and rectal injury the lowest value. The results of the corresponding cost-effectiveness model were found to be sensitive to changes in these utility values, showing that the decision whether or not to screen for prostate cancer is dependent on the preferences of men.

As the authors point out, these results were based upon preferences of men in a primary care setting. The results may not be representative of men in specialty clinics who have either been referred by their physician or self-referred, and who therefore may be at a higher risk of prostate cancer. It is also the case that the sample was very small.

Volk and colleagues[275]

Volk and colleagues followed up the Cantor study,[274] using the same study sample, but also interviewing the wives of the 10 men originally interviewed. Utilities of husbands and wives were elicited separately, and then elicited together as a couple. The cost-effectiveness model showed that the optimal screening strategy in terms of maximising quality-adjusted life expectancy was no screening for seven out of 10 husbands, compared with a screening strategy for nine out of 10 of the wives. The wives appeared to wish to maximise quantity of life for their husbands, paying less attention to quality than the husbands paid on their own account. In fact, utilities elicited from wives on behalf of their husbands were consistently higher than the utilities elicited from the men except for thromboembolisms, for which the mean male utilities were higher. The utilities from the couples were between those for the men and their wives. When the 'couple' utilities were fitted, the results were the same as for the husbands' utilities.

Like the study by Cantor and colleagues, this study did not take into account the possible short-term psychological aspects of screening, such as anxiety or regret, and it suffers from a very small sample size.

Krahn and colleagues[268]

Krahn and colleagues considered the cost per QALY of one-off prostate cancer screening, including utilities elicited from 10 physicians using TTO. This was a very small physician sample and the effects of the screening process itself were not considered.

Johnston and colleagues[222]

Johnston and colleagues constructed five descriptions of health states: good health (as for a woman of a similar age to the interviewee), true negative, false positive, true positive and false negative. A total of 440 women aged 40–64 years were interviewed. Health states were described in bullet format and described receipt of the invitation, the waiting time between the invitation and the appointment, the process of screening, and also quality of life associated with breast cancer treatment for the true-positive and false-negative outcomes. These descriptions were based on the findings of a literature review. The TTO and the VAS methods were used for comparison. Women were also asked to rank the health states before valuing them. Interestingly, 9.8% ranked false positive as preferable to true negative, with some women indicating that they would feel additional reassurance that they were indeed cancer free from the extra tests conducted due to the false-positive result. Some 7.6% of the sample ranked the false-negative result above the true-positive result, although it may be that these women would not attend for screening.

The findings of the study were varied according to the method used. According to the TTO results, the differences between the two permanent states (true positive and false negative) were not statistically significant, whereas there were significant differences between the two temporary states (true negative and false positive) according to the *t*-test and Wilcoxon ($p < 0.05$). The VAS results produced some different results, such as the false-positive state being given a lower value than the true-positive state according to both mean and median values. The true-negative and false-positive states and the true-positive and false-negative states were significantly different for both means and medians ($p < 0.05$).

Johnston and colleagues put forward a number of possible weaknesses in their study. The relatively

large difference between the true-negative and false-positive results may in part be due to the use of alternative anchor points between the two temporary states. There appeared to have been some possibility of cognitive overload in the valuation processes, particularly for the VAS method, to which some respondents gave values of worse than death to one or more of the screening outcomes (including the true-negative one).

Gerard and colleagues[279]
Gerard and colleagues extended the study by Johnston and colleagues,[222] using the same sample, to map the TTO scores for the screening outcomes to the EQ-5D scale. The majority of the sample (75.7%) judged the true-negative state to be equal to the EQ-5D state of full health (11111), although the true-negative state was mapped on to a total of 12 EQ-5D states by different individuals. The false-positive state was mapped on to 18 EQ-5D states. The true-positive state was mapped to 20 EQ-5D states and the false-negative state to 19 EQ-5D states.

This study shows that different individuals within the sample frequently interpreted the condition-specific breast screening outcome descriptions differently. This must raise some concerns about the accuracy of the descriptions in the vignettes. It would have been better to have administered the EQ-5D to patients in the relevant states which would have shown the range of experiences captured in the simple vignettes. The TTO, VAS and EQ-5D data obtained from Johnston and colleagues[222] and Gerard and colleagues[279] also show how different results can be obtained from different valuation methods, but this is not a new discovery.

Essink-Bot and colleagues[270]
The study by Essink-Bot and colleagues took a different approach to that used by the studies described above. Rather than using utility elicitation tools to value hypothetical health states, they used SF-36 and EQ-5D descriptive systems to obtain health status and HRQoL values for men at multiple points between receiving an invitation to prostate cancer screening and being informed of their screening results. The authors also included the State–Trait Anxiety Inventory (STAI), which is a 40-item questionnaire designed to measure anxiety due to specific situations and also the degree to which an individual is anxiety-prone by nature. The authors hoped to use the results to determine the extent to which participants of screening for prostate cancer were affected in the short term by the screening test itself or a biopsy if

this was required. The sample size at the beginning of the study was 600 screened men and 235 non-participants of screening. The times at which screened men were surveyed were designated as follows:

- T1 (3 weeks prior to screening, SF-36 and EQ-5D data obtained)
- T2 (in the waiting room prior to screening, STAI and EQ-5D data obtained)
- T3 (1 week after being notified of a negative result, STAI and EQ-5D data obtained)
- T4 (during the 2-week period between biopsy and receipt of the result, STAI and EQ-5D data obtained)
- T5 (1 week after receiving a clear result from the biopsy, SF-36, EQ-5D and STAI data obtained).

Small but significant improvements were observed in comparisons of SF-36 scores at T3 and T1 for true-negative results and at T5 and T1 for false-positive results. No significant differences were found between the EQ-5D scores tested between time periods T3:T1, T3:T2, T5:T1, T4:T2 and T5:T4. Comparisons between STAI scores indicated that mean anxiety scores improved between receiving the true-negative results and awaiting screening (T2), and there was also an improvement between receiving the clear result from the biopsy and the period while awaiting the result of the biopsy.

From these results, it appears that, according to the EQ-5D, the screening process itself, and whether participants received a true-negative or a false-positive result, made no difference to HRQoL. This was despite a considerable proportion of respondents reporting pain and/or discomfort during the screening and biopsy procedures.

This study would seem to provide the most conclusive evidence to date that the screening process does impact on HRQoL, although whether the overall impact is positive cannot be understood from this study alone since it does not examine all possible screening outcomes and its T1 already incorporates an element of prescreening anxiety. This study has the advantage that it actually records the way people responded to the experiences during the process rather than people trying to imagine what it is like.

Cormier and colleagues[276]
Cormier and colleagues also studied the effects on HRQoL and anxiety of screening for prostate

cancer. They asked a sample of 220 sons or brothers of men with prostate cancer, who were undergoing familial prostate cancer screening, to complete SF-36 and STAI questionnaires prior to screening, after screening, and after receiving negative results.

The statistical analyses used both linear and quadratic changes in HRQoL over time. There were significant increases in HRQoL over the dimensions of general health, vitality and mental health for the linear model ($p < 0.05$), and the quadratic model noted a significant increase in the score for role emotional ($p < 0.05$). The STAI score improved over time according to both the linear and quadratic models ($p < 0.05$), implying that members of the sample grew less anxious from the beginning to the end of the screening process. The authors determined that "moderate anxiety deterioration and minimal HRQoL deterioration" occurred in approximately 20% of the sample during the whole screening process. Factors associated with deterioration in HRQoL included being aged between 50 and 60 years, more than two relatives with prostate cancer, a high-level anxiety trait, a high educational level and no children currently living at home.

The study shows that over the screening process there are some small and yet significant changes in HRQoL. However, again the baseline measurements may incorporate some screening anxiety so the overall impact of the screening is not known.

Edelman and colleagues[277]
Edelman and colleagues examined the effects of screening for diabetes on HRQoL 1 year after the screening process, looking particularly at the effect of labelling people. A total of 1253 individuals without known diabetes were screened by a blood test, 94% of whom were men; 56 individuals were screened positive for diabetes. Scores were lower 1 year later for those screening positive and negative. It is not possible to say that the decline was due to screening, as the sample had aged a year and many had illnesses other than diabetes, and a multivariate analysis indicated that co-morbidity was the major predictor of HRQoL. The impact of false-positive or false-negative results is not discussed, and since this study did not collect SF-36 data closer to the time of screening than 1 year later, it is unknown whether screening may have had short-term effects that could have disappeared by the time the data was collected after 1 year.

Hensley and colleagues[271]
Hensley and colleagues compared pre- and post-menopausal women who were at high risk of developing ovarian cancer in terms of HRQoL, cancer-related anxiety, depression and perceptions of their risks of developing ovarian cancer. A total of 147 high-risk women who were being screened twice annually were asked to complete the SF-36, the Impact of Events Scale (a measure of cancer-related anxiety), and a measure for depression from the Centre for Epidemiologic Studies. Respondents were also asked to rate their perception of their lifetime risk of ovarian cancer compared with other women of a similar age and family history, and also to rate their perceptions of their lifetime risk on a scale from 0 to 100. They completed these assessments at each screening attendance.

The authors did not provide much detail around the SF-36 results, just stating that "general health-related quality-of-life scores on the SF-36 ranged from 61 to 88, indicating generally good quality of life among cohort participants. There were no differences in quality-of-life scores between pre- and postmenopausal women". The authors commented that the SF-36 scores were comparable to the general population. It would have been very interesting to have more information on the SF-36 as it may have provided information on HRQoL during the screening process.

Rumbold and Crowther[273]
Rumbold and Crowther conducted a prospective study of women being automatically screened for gestational diabetes mellitus (GDM) in Adelaide, Australia. They collected SF-36 data, data on anxiety using a six-item short-form of the STAI,[282] and data on depression using the Edinburgh Postnatal Depression Scale.[283] Women were also asked about their perception of health, their concerns for the health of their baby, adequacy of information about the screening test and diagnostic test and their overall experience of being screened.

A cohort of women were followed through pregnancy, and were questioned prior to screening, after screening and after receipt of results in early pregnancy, and in late pregnancy. Some women who gave consent after screening positive were also included. Although women screening positive were more likely to rate the screening experience negatively, there was no difference in choosing whether to be screened again in a future pregnancy, with approximately half of both groups saying they would. The

majority of women thought the tests had been explained adequately.

There were no significant differences in STAI or depression scores between women screened positive and negative, and there were no differences in concern for the health of the baby between screening groups. After screening in early pregnancy, those screened positive scored significantly higher on vitality, but lower on general health perceptions. Screened negative women were more likely to rate their health as "much better than 1 year ago". In late pregnancy, those with negative tests scored significantly higher for social functioning than those who received positive or false-positive test results. The authors felt that the fact that women screening positive were more likely to rate their health less well was not so much because they were unhealthier, but more due to their perceptions being altered as a result of the screening test. The findings of this study were that screening had a negative impact on health perceptions in women screening positive, although the degree to which this matters was not discussed.

Fry and colleagues[272]

Fry and colleagues conducted a postal retrospective study to compare 27 women who had elected to have an oophorectomy 1–5 years previously due to high familial risk with 28 high-risk women who had elected to undergo ovarian screening. The authors collected SF-36 data, information on menopausal symptoms, the General Health Questionnaire, a questionnaire to assess cancer-related worry, a questionnaire to assess sexual adjustment and a questionnaire on body image and gender identity. Women who had undergone oophorectomy rated their mental health, role emotional, social functioning and bodily pain significantly lower on the SF-36 than those who had elected to be screened. The role emotional and social functioning scores remained significantly lower after removal of those in the sample who had previously suffered from breast cancer. The authors concluded that a larger study was required, and their sample was indeed very small.

Shackley and Cairns[269]

Shackley and Cairns were interested in exploring the issue of screening in terms of the value of information gained rather than merely the values attached to screening health states. They chose the area of antenatal screening for cystic fibrosis (CF) and undertook interviews with 52 women aged between 18 and 45 years, with no current

pregnancy, and no knowledge of CF-carrying status. During the interview, each respondent answered six SG questions. For each question each woman was handed a card which described the risk that her hypothetical baby would be born with CF. She was told that an amniocentesis would inform her as to whether or not her child had CF. A series of cards describing risks associated with amniocentesis were given to her to read, and the level of risk (p) at which she was indifferent between the risk of abortion and the indeterminate screening probability. From this it was possible to calculate the utility of the information ($1 − p$).

Applying this utility information to populate a screening decision tree, screening was preferred to not screening by 69% of the sample, with 29% preferring not to be screened and 2% being indifferent. The median expected utility value of screening was 0.999755 and of not screening 0.999000. Although this is a very small difference, it proved to be highly statistically significant using the Wilcoxon matched-pairs signed-ranks test ($p = 0.0004$).

Cairns and colleagues[278]

Cairns and colleagues describe further data from the sample used by Shackley and Cairns.[269] They used the utility values obtained to calculate utilities for two types of screening programme for CF: stepwise screening and couple screening. The median expected utilities for stepwise screening, couple screening and no screening for the group overall were 0.99975, 0.99972 and 0.999 respectively. The preferred screening strategy depended on whether all women were screened. If all women (including those who preferred no screening) were screened, the stepwise screening would yield the highest expected utility. However, if only those women for whom screening is preferred are considered, couple screening is the preferred strategy. Shackley and Cairns[269] and Cairns and colleagues[278] made the assumption that the order of preference is full information about the CF status of the fetus, partial information and finally abortion brought on by amniocentesis. However, since it is possible that some women would prefer not to have full information, the study ought to have begun by asking women to rank these scenarios.

Hall and colleagues[234]

Hall used healthy-year equivalents (HYEs) to measure utilities for breast cancer scenarios to assess the cost–utility of a mammography screening programme in Australia, arguing that

existing generic preference-based measures were not relevant to the outcomes being considered. They used a single-stage TTO to assess HYEs in 104 women, 44 of whom served as a community sample and the remaining 60 of whom had suffered from breast cancer. Scenarios comprised type of surgery (radical or conservative), physical health (good or poor) and mental health (good or poor). Scenarios ending in death from breast cancer were valued significantly lower than those ending in death from unspecified cause, but type of surgery did not appear to have a significant effect on utility. Hall assumed that utility would not depend on whether cancers were screen-detected or not, and therefore did not value the screening process itself. The study focused on true positives.

Discussion

The studies described in the previous section differed widely. The size of the sample informing different studies varied greatly. Indeed, some of the small sample studies may be considered pilot studies. Some studies obtained utilities from health professionals, whereas others elicited utilities from screening participants themselves. There is evidence that clinicians give different values to health states from those given by patients, often rating patient health higher than the patients would rate their own health.[284] NICE requires general population values for assessing cost-effectiveness for their reference case and this may limit the usefulness of some of the data. Other researchers used members of the general population.

A wide variety of methods were used to elicit preferences. Whereas some studies used traditional health economic evaluation techniques such as SG, TTO and VAS to obtain utilities for hypothetical states (e.g. Shackley and Cairns[269]), other studies used generic HRQoL measures such as SF-36 to value current health at different points in the screening process (e.g. Rumbold and Crowther[273]). The EQ-5D has also been used (e.g. Essink-Bot and colleagues[270]), which is valued using the TTO. As Johnston and colleagues[222] and Gerard and colleagues[279] showed, different methods of valuing the same data (VAS, TTO and values mapped to EQ-5D) produced different results. Hall and colleagues[234] used a TTO-based version of the HYE to explore the utility of hypothetical health profiles for mammography screening. This variation in methods makes it difficult to compare across studies.

Some of the studies (e.g. Essink-Bot and colleagues[270]) looked at the effects of screening in the short term, whereas other studies (e.g. Edelman and colleagues[277]) looked at the effects in the longer term. Although longitudinal studies are better than cross-sectional studies (where differences are difficult to interpret), there remains the problem that at baseline the population prior to screening has already been alerted to the problem and so may have artificially lowered scores.

The use of vignettes has the attraction that it means the researcher is able to make the content more relevant to the screening state. However, there are real concerns that these vignettes may not be representative of the variance of experience that would be reflected in a longitudinal study of patients undergoing screening. However, those studies that follow-up patients through the screening process have tended to rely on generic measures of health status since only these provide the preference-based scores required for economic evaluation. However, it is unclear whether the SF-36 and EQ-5D instruments are sensitive enough to detect changes in HRQoL specific to screening. As Hall[234] notes, generic measures do not always capture the types of changes in HRQoL that screening might produce. Hall was referring specifically to the Rosser scale, although similar might be said of the SF-36 and EQ-5D. The EQ-5D asks about the five dimensions of mobility, self-care, usual activities, pain or discomfort, anxiety or depression. The first three are unlikely to be affected by the screening process, although it is possible that screening could cause some pain or discomfort, or cause anxiety or depression. However, there may be disutility associated with screening that would not fit into these classifications.

The findings from the longitudinal studies were difficult to interpret. Essink-Bot and colleagues[270] found that, according to SF-36 scores, HRQoL improved after screening for prostate cancer, and this was after both true-negative and false-positive results. However, Rumbold and Crowther[273] found that pregnant women being screened for gestational diabetes gave higher scores for some of the SF-36 dimensions if screened negative than false positive. As argued earlier, the baseline assessment may incorporate the prospect of screening.

This review has identified considerable gaps that exist in the evidence on screening outcomes. One of the reasons for this is the difficulty of

undertaking research in this area. The vignettes approach suffers from the necessary trade-off between the adequate description of the range of consequences of different actions and the dangers of cognitive overload, whereas longitudinal studies tend to be already contaminated by the very existence of a screening programme. Well-designed studies should be able to obtain descriptive quality-of-life data describing the consequences of different screening effects, and generic preference-based measures could be used, with the SF-36 seeming to be better able to detect in true-negative and false-positive cases,[270] or perhaps one based on a more specific descriptive system.

Screening outcomes differ considerably by programme, and so the research would have to be specific to disease and programme. Such studies need to be large in scale and properly controlled (for example, obtaining baseline health state utility values without the immediate prospect of screening). The extent to which such research is required for economic models needs to be determined through rigorous value of information analysis.

There are other methods of valuing the benefits of screening, such as willingness to pay or discrete choice experiments, though this review was restricted to evidence for use in cost-effectiveness models, and hence to studies using recognised health state valuation methods or preference-based measures (plus the SF-36 due to the possibility of deriving a preference-based index).

Conclusion

Existing studies provide little data for populating economic models. Few of the studies looked at the whole range of screening outcomes. Since the studies varied in methods, samples and screening programmes, it is difficult to pool the results of different studies that looked at different aspects of screening utilities. Much of the previous research has used sample sizes that were too small for the results to be generalisable. They either used vignettes that are hard to validate in the absence of longitudinal data or they were longitudinal studies suffering from having baseline values that are already contaminated by the prospect of screening.

However, the evidence does suggest that the psychological consequences of true-negative and false-positive results can have statistically significant effects on HRQoL in some cases. These effects seem to be modest at the individual level, although over a population they may be important. Also at an individual level, there are a number of important side-effects and complications from diagnosis and treatment that would need to be counted in any assessment of the costs and benefits of screening. Where value of information analysis finds the value of one or more of these outcomes to be sufficiently important for a given proposed screening programme, then properly controlled large-scale longitudinal studies should be undertaken using either a generic preference-based measure or a more specific preference-based measure developed for the task.

Chapter 10
Guidelines and good practice for model-based cost–utility analyses of screening programmes

Introduction

These guidelines are designed to assist modellers in ensuring the quality of model-based cost–utility analyses of healthcare screening programmes. There are various guidelines informing the conduct of decision analytic modelling in health technology assessment,[2–4] which have recently been reviewed to produce a synthesised set of guidelines and an accompanying checklist.[5] The screening modelling guidelines presented in this chapter focus on issues that are specific to the modelling of screening programmes.

The guidelines have been defined by the authors of this report, and are informed by the findings from the review of screening models, the application of the reported case study evaluations, and discussions over the course of the review with health economists and operational researchers with experience of modelling screening programmes (as detailed in the Acknowledgements).

The guidelines

The guidelines are presented under the following broad categories:

1. research question
2. general modelling approach
3. model structure
4. modelling technique
5. model population
6. validation and calibration
7. issues specific to antenatal screening.

The research question
The research question defines the boundaries of the modelling approach.

1.1. Cost–utility analysis of the lifetime costs and QALYs is appropriate for most screening programmes.

1.2. Practicalities of implementation should be clarified prior to defining the research question, for example, if repeat (incident) screens, non-constant screening intervals or varying screening policy by birth cohort are not feasible, then such options should be excluded from the research question.

The general modelling approach
Natural history modelling is the preferred general approach. These models describe disease progression from the point at which disease becomes detectable to death. When a screening model is laid on top of the natural history model, the course of the natural history is altered through the detection of disease at an earlier stage.

The main advantage of the natural history approach is that it permits the evaluation of screening options that have not been observed in primary research. Only estimates of test sensitivity and specificity are required. The non-natural history models included in the review tended to have less intuitive pathways due to difficulties incorporating the timing and frequency of screening and related activities. The main disadvantage of the natural history approach is the input data requirements, which usually include various parameters that are not directly observable. These issues are addressed in the section 'Model population' (p. 119).

2.1. The evaluation of a prevalence (one-time) screening programme is generally simpler than the evaluation of incidence (repeat) screening programmes as the model is only concerned with disease that is detectable at a specific age. If prior opinion indicates that any form of screening is unlikely to be cost-effective, a simple prevalence screening model (e.g. a decision tree model) incorporating assumptions that favour no screening may be used as a hurdle to the development of a more complex modelling evaluation.

2.2. Models that extrapolate observed intermediate end-points to estimate lifetime costs and QALYs may be used when all relevant screening programmes have been observed.

2.3. Natural history models should be used to evaluate the lifetime costs and QALYs of repeat

screening programmes when one or more of the relevant screening programmes have not been directly observed.

The model structure

The ISPOR Task Force[2] states that "the model structure should be consistent both with a coherent theory of the health condition being modelled and with available evidence regarding causal linkages between variables" (p. 11) and that "[t]he structure of the model should be as simple as possible, while capturing underlying essentials of the disease process and interventions" (p. 12).

More complex model structures may incorporate important additional aspects of the disease natural history, although any benefits should outweigh the consequences of additional unobservable input parameters and increased complexity in implementing the model. No direct comparisons of more detailed and less detailed screening model structures were identified to inform empirically areas in which more realistic representations of the disease process may be most beneficial, so the following points relate to general issues of good practice.

3.1. The process of defining a model structure is an iterative process that involves a review of existing screening models; feedback to clinical experts; expert guided review of the natural history literature; feedback to experts with draft model structure; and discussions between clinical experts and informed analyst about the strength of assumptions, calibration requirements and the feasibility of model development.

3.2. To assist the explicit process of model development, a complex model structure should be initially defined. The iterative process can inform areas in which simplifying assumptions that reduce model complexity can be justified.

3.3. Disease states at the point of clinical presentation should represent prognostic indicators that influence treatment choices and treatment effectiveness; for example, a breast cancer screening model may describe health states as a function of tumour size, nodal status, oestrogen receptor status and menopausal status. If discrete states are used, the categorisation should reflect the relationship between the prognostic indicator and treatment choices and treatment effectiveness.

3.4. The categorisation of disease states may be informed by data available to populate the model,

or data describing output parameters against which the model can be calibrated.

3.5. Whynes and colleagues presented a linear regression analysis investigating the impact of stage at diagnosis on age at death, and a Cox regression analysis determining the hazard ratio by participation in screening.[132] Where possible, similar analyses should be undertaken to assess the sensitivity of a model's disease categorisation to differences in prognosis between screen-detected and clinically presenting patients.

3.6. The post-diagnosis section of the model should describe disease progression to death using treatment models that are representative of current treatment patterns for different stages of the disease. There are two broad approaches to integrating treatment models with screening models:

- Separate treatment models are defined for each separate clinical presentation state.
- A single treatment model describes the possible progression of patients from the earliest point of clinical presentation through subsequent stages to death, in addition to enabling patients to enter the model at each subsequent stage.

3.7. Evidence suggests that screening attenders may have different characteristics to non-attenders,[270] which may influence some disease parameters (e.g. disease incidence or progression).[137] The model structure should incorporate these effects, either by:

- assigning individuals characteristics that influence screening uptake and disease parameters, or
- facilitating alternative disease parameters for attenders and non-attenders.

The modelling technique

The most commonly applied modelling techniques to describe lifetime costs and effects (either life-years or QALYs gained) are cohort Markov models and individual sampling simulation models (either Markov models or DES). More recently, some complex mathematical models have analysed the cost effectiveness of screening.

Cohort Markov models provide a straightforward approach to describing relatively simple model structures that remain within the bounds of the Markovian (no memory) assumption. Additional flexibility can be incorporated through the inclusion of additional health states, although

their feasible application may be limited by the expansion in the number of health states required. The following are factors that increase the number of health states included in a cohort Markov screening model:

- time-in-state dependent transition probabilities
- transition probabilities influenced by past events
- transition probabilities influenced by individual attributes
- separate disease components within individuals.

Individual sampling Markov models and DES can overcome the above limitations of the cohort approach by storing relevant characteristics and aspects of individuals' pathways as attributes that influence future pathways, for example, the number of years that an individual remains in a particular state may determine the transition probability applied in the next cycle. Individual sampling models can have significantly longer running times, especially when probabilistic sensitivity analyses are required.

4.1. The choice of modelling technique is largely dependent on the defined model structure. Individual sampling simulation models may provide more flexibility in their representation of a screening decision problem, but any benefits should outweigh the consequences of the need to assess variability in addition to uncertainty.

4.2. The modelling techniques required to implement alternative model structures should not influence the defined model structure, although in practice resource constraints (time and analytic expertise) may inform some structural choices.

4.3. More complex mathematical models that describe input parameters as continuous variables require further development to estimate the cost–utility of screening directly, or to inform a more detailed representation of the preclinical section of a natural history model (with a traditional state-based model describing pathways post-clinical presentation).

4.4. Another potential modelling option is mixtures of Markov models, which addresses heterogeneity within a population by assigning individuals to different clusters, where each cluster has a different Markov model. As a patient enters the eligible population for a screening programme, they are assigned to a particular cluster with some probability and the pathway for that individual is then generated from a Markov model with parameters specific to that cluster.

Model population
The sections below consider a range of input parameters that are of particular relevance to screening models.

5.1. In the absence of longitudinal observational studies, age-specific incidence rates and progression rates for preclinical disease should be calibrated to age-specific prevalence rates of different stages of preclinical disease (informed by population-based studies using a gold standard diagnostic test, or by autopsy studies). Incidence rates based on the assumption of linear progression between prevalence rates at different ages fail to account for disease progression in the interim period.

5.2. To reduce the number of unobserved parameters, time-varying probability distributions, such as the Weibull distribution, should be considered rather than the estimation of a separate incidence rate for each age group.

5.3. The most appropriate methodology to estimate sensitivity and specificity of relevant screening tests will vary according to characteristics of the disease and the screening study. The disease lead time affects the accuracy of the work-up and follow-up method, and selection procedures for diagnostic evaluation may vary by screening study. A suggested hierarchy of alternative study designs is as follows:

1. population-based studies comparing screening test(s) with a reference standard test
2. 'screening' before and independently from diagnostic evaluation
3. 'screening' patients with known disease
4. diagnostic work-up of test-positive individuals and the follow-up of test-negative individuals to identify missed cases
5. diagnostic work-up of test-positive individuals and test-negative individuals, and the application of a work-up bias correction method
6. synthesis of the observed (age-specific) screen-detection rate with (age-specific) estimates of prevalence.

5.4. Methods for pooling multiple estimates of screening test characteristics include the summary receiver operating characteristic (ROC) approach, and bivariate models of sensitivity and specificity. The summary ROC approach can be modified to obtain correlated estimates of sensitivity and specificity, but requires complex Markov chain Monte Carlo methods. The bivariate model

approach provides an additional ability to incorporate explanatory variables and can now be analysed using standard statistical packages.[285]

5.5. Input parameters for post-diagnosis disease progression, costs and utility values should reflect current treatment options. Existing models used to evaluate interventions at different stages post-diagnosis can be linked to the natural history model.

5.6. The screening model should incorporate the fixed and variable costs of screening programmes, so that the fixed cost is applied regardless of uptake, while the variable cost is applied to each screened individual.

5.7. Data describing screening uptake rates by screening test and screening round should be sought.

5.8. Evidence of differential disease prevalence and progression rates in populations accepting and declining the offer of screening should be sought.

5.9. Evidence of utility effects specific to the screening tests under evaluation should be sought and incorporated into the model, at a minimum, in the sensitivity analysis.

Validation and calibration

Validation assesses the accuracy of a model by comparing the outputs of a fully populated model to observed data. Calibration involves fitting a model to observed data describing outputs of the model in order to populate unobserved input parameters. All natural history screening models contain unobservable input parameters, although the review identified studies that estimated values for unobservable parameters using extrapolations of observed growth rates, expert opinion or previously developed theories.

6.1. Data-based estimation of parameter values is generally preferred, although the advantages of this approach are only realised when the model is validated.

6.2. Output parameters include age- and stage-specific incidence and prevalence of clinical disease and disease-specific mortality rates. Consideration should be given to the comparability of the output data obtained. The relevance of alternative data sources can be reflected by weighting the output data.

6.3. The estimation of a reference case input parameter set is not recommended. A preferred calibration process involves predicting output

parameters for a large number of input parameter sets, with the accuracy of each set's predictions represented as a weight. If multiple outputs are predicted, a weighted sum of the weights across all the outputs is used to describe the likelihood that each input parameter set is the most accurate set of input parameters. The main analysis of the model involves sampling a large number of input parameter sets according to the weights attached to each input parameter set, from which mean values and probability distributions of cost-effectiveness can be derived.

6.4. Cross-validation with other screening models should be undertaken. The validity of a model will be increased if variations in the outputs of alternative models can be explained by differences in the assumptions and input parameter values.

Issues specific to antenatal screening

7.1. Outcomes of antenatal screening programmes in which the termination of affected pregnancies is an option include the number of cases detected, informed choices provided and affected pregnancies terminated. It is also important to describe the number of false-positive screens that result in screening-related miscarriages and terminations of unaffected pregnancies. The complex moral issues involved mean that the estimation of QALYs is inappropriate for such screening evaluations.[138] The relevance of estimating the long-term cost impact of terminating affected pregnancies should be discussed with the policy maker, where appropriate. Otherwise, such costs should be presented separately for the potential consideration by the user.

7.2. The estimation of the lifetime costs and QALYs is appropriate for the evaluation of antenatal screening in which the early detection of a condition is the primary outcome of interest (e.g. antenatal HIV screening). The large uncertainties around predicting lifetime effects for newborns should be incorporated in the analysis.

7.3. The inclusion of time dependent input parameters in antenatal screening models should be considered, particularly in relation to the uptake of prenatal diagnosis and terminations of pregnancy.

7.4. Decision trees should be used to model simple antenatal screening pathways due to their high level of transparency. Adapted decision trees incorporating a partial programming approach or fully programmed models are similarly effective for representing more complex antenatal screening model structures.

Chapter 11
Assessment checklist

Introduction

The checklist is designed to assist users and reviewers to assess the quality and relevance of model-based cost–utility analyses of healthcare screening programmes. The assessment checklist is informed by the guidelines and good practice reported in Chapter 10, which were examined to identify a set of issues against which the quality of a screening model could be assessed.

The checklist

The checklist covers the same categories as used to define the guidelines in Chapter 10:

1. research question
2. general modelling approach
3. model structure
4. modelling technique
5. model population
6. validation and calibration
7. issues specific to antenatal screening.

The research question
1.1. If not evaluating antenatal screening, does the model estimate the lifetime costs and QALYs of alternative screening programmes? If not, what justifications are presented?

1.2. Does the research question include the evaluation of all relevant screening options, in particular, repeat (incident) screens? If not, what justifications are presented?

The general modelling approach
2.1. Is a natural history model used to describe disease progression from preclinical incidence to death, with alternative screening programmes overlaid? If not, is this because only observed screening programmes, or only prevalence (one-time) screening programmes are evaluated?

The model structure
3.1. Is the process of defining the model structure explicitly described?

3.2. Are any simplifying assumptions that reduce model complexity justified?

3.3. Are the prognostic indicators by which disease states are defined justified in terms of their impact on treatment choices and treatment effectiveness?

3.4. Is the categorisation of the prognostic indicators by which disease states are defined justified? Is the sensitivity of the disease categorisation to differences in prognosis between screen-detected and clinically presenting patients analysed and/or discussed?

3.5. Does the post-diagnosis section of the model describe disease progression to death using treatment models that are representative of current treatment patterns for different stages of the disease?

3.6. Does the model facilitate the alternative representation of disease incidence and progression for screening attenders and non-attenders?

The modelling technique
4.1. Is the choice of modelling technique justified? If an individual sampling simulation model is used, are the reasons for not using a cohort Markov model (i.e. areas of increased complexity) described?

4.2. If more complex mathematical modelling approaches are used, do they adequately represent the cost and utility effects of alternative health states?

Model population
5.1. Are preclinical disease incidence and progression rates appropriately estimated, either using longitudinal observational studies or calibrated to age-specific prevalence rates, and not estimated by experts or based on the assumption of linear progression between prevalence rates at different ages?

5.2. Are the numbers of unobserved parameters for the specified model structure minimised, for example, are time-varying probability distributions used rather than separate incidence rates for each age group?

5.3. Are inclusion criteria for studies informing screening test sensitivity and specificity rates

justified on the basis of the characteristics of the disease (e.g. disease lead time affects the accuracy of the work-up and follow-up method), and the conduct of alternative study types (e.g. selection procedures for diagnostic evaluation may vary by screening study)?

5.4. Do the methods for pooling multiple estimates of screening test characteristics control for heterogeneity between studies?

5.5. Do input parameters for post-diagnosis disease progression, costs and utility values reflect current treatment options?

5.6. Are the fixed and variable costs of screening programmes incorporated?

5.7. Are screening uptake rates described as a function of screening test and screening round?

5.8. Was evidence of alternative disease prevalence and progression rates in populations accepting and declining the offer of screening sought? If found, was it incorporated in the model?

5.9. Were utility effects specific to the disease and screening tests under evaluation sought and incorporated in the model?

Validation and calibration
6.1. If the model was not calibrated, was it validated?

6.2. Are the output parameters used to calibrate or validate the model relevant to the perspective of the evaluation?

6.3. Are the number of output parameters and the quality of the available data informing their estimation discussed in relation to the calibration of the model?

6.4. Does the main analysis of the model sample a large number of input parameter sets according to the weights attached to each input parameter set, from which mean values and probability distributions of cost-effectiveness are derived?

6.5. Are the results of the model compared with those of other screening models, and differences in the outputs of alternative models explained by differences in the assumptions and input parameter values?

Issues specific to antenatal screening
7.1. If evaluating antenatal screening, does the model describe the range of important outcomes, including cases detected, affected pregnancies terminated, screening-related miscarriages and terminations of unaffected pregnancies?

7.2. If the early detection of a condition is the primary outcome of interest, does the model estimate lifetime costs and QALYs? If not, is the justification around the large uncertainties around predicting lifetime effects for newborns?

7.3. Does the model include time-dependent input parameters (e.g. in relation to the uptake of prenatal diagnosis and terminations of pregnancy)? If not, is the exclusion justified?

7.4. Is the choice of modelling technique justified? If adapted decision trees incorporating a partial programming approach or fully programmed models are used, are the reasons for not using a simple decision tree (i.e. areas of increased complexity) described?

Chapter 12
Conclusions

The review of methods for the model-based cost–utility analysis of screening programmes identified the natural history modelling approach as the preferred general method of evaluation for screening programmes. State transition models have generally been used to represent disease natural histories, with individual sampling models more prevalent than in treatment intervention evaluations. No comparative methodological studies were identified, so no empirical data were available to inform the relative merits of alternative methodologies. The defined guidelines and assessment checklist are informed, therefore, by theoretical interpretations of the impact of alternative approaches to different components of the modelling process when applied to the cost–utility analysis of screening programmes.

Recommendations for further research

Areas for further research are listed below. They are based on methods with the potential to improve the accuracy of screening models, and to respond to the needs of model users.

More complex mathematical modelling approaches have great potential as an alternative or adjunct to state-based modelling techniques for the evaluation of the cost–utility of screening programmes. Research is needed into the development of such models for the full evaluation of the cost-effectiveness of screening, in addition to a hybrid formation in which such techniques may be best suited to modelling the preclinical phases of disease.

There is scope for developing more comprehensive and explicit methods for calibrating models, which describe correlations between input parameters.

Empirical estimates of differences in the mean and probabilistic outputs of less complex cohort Markov models and more complex individual sampling models, using the same data sources, would be of interest. Such comparisons may inform general areas in which simplifying assumptions are justified.

The direct utility effects of screening are under-researched, and may have a significant effect on the estimated cost–utility ratios. More primary screening studies should incorporate utility measurements in their protocol.

The review of screening models has identified that more complex models may have greater relative benefits in the area of screening than in other areas of evaluation. Further research is needed into the extent to which policy makers view transparency as an important characteristic of screening models, and whether existing methods of verification and validation are sufficient.

Acknowledgements

Contribution of authors

Jon Karnon (Senior Research Fellow) undertook searches, abstract review and review of full papers retrieved, wrote Chapters 1, 2, 6, 7 and 10–12, co-wrote Chapters 3, 4 and 8, edited the remaining chapters and combined them into the final report. Elizabeth Goyder (Clinical Senior Lecturer in Public Health Medicine) undertook the abstract review and wrote Chapter 5. Paul Tappenden (Research Fellow) co-wrote Chapter 4, Seonaid McPhie (MSc placement student) co-wrote Chapter 8, Isabel Towers (Research Fellow) co-wrote Chapter 9, John Brazier (Professor of Health Economics) co-wrote Chapter 9 and Jason Madan (Research Fellow) co-wrote Chapter 3.

◆

References

1. Wilson JMC, Junger YG. *Principles and practice of screening for disease*. Geneva: World Health Organization; 1968.

2. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, *et al*. 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* 2003;**6**:9–17.

3. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *PharmacoEconomics* 2000;**17**:461–77.

4. Gold MR, Siegel E, Russell LB, Weinstein MC. *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.

5. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36).

6. Boyd A, Davies LA, Bagust A. Modelling the implications for hospital services of cervical cytology screening: a case history. *J Oper Res Soc* 1989;**40**:529–37.

7. Laporte G, Semet F, Dadeshidze VV, Olsson LJ. A tiling and routing heuristic for the screening of cytological samples. *J Oper Res Soc* 1998;**49**:1233–8.

8. Laporte G, Brotcorne L, Semet F. Fast heuristics for large scale covering-location problems. *Comput Oper Res* 2002;**29**:651–65.

9. Benneyan JC, Kaminsky FC. Statistical and economic models for analysis and optimal design of laboratory screening policies for cervical cancer. *Ann Oper Res* 1996;**67**:235–85.

10. Baker RD. Use of a mathematical model to evaluate breast cancer screening policy. *Health Care Manage Sci* 1998;**1**:101–13.

11. Baker RD. A methodology for sensitivity analysis of models fitted to data using statistical methods. *IMA Journal of Management Mathematics* 2001;**12**:23–39.

12. Baker RD. What can industrial inspection models learn from medical analogues? Determining the optimum screening policy for breast cancer. *IMA J Math Appl Bus Ind* 1998;**9**:305–24.

13. Baker RD. How to correctly calculate discounted healthcare costs and benefits. *J Oper Res Soc* 2000; **51**:863–8.

14. Baker RD. Sensitivity analysis for healthcare models fitted to data by statistical methods. *Health Care Manage Sci* 2002;**5**:275–81.

15. Dixon M. *ABC of breast diseases*. London: BMJ Books; 1995.

16. Govan ADT, MacFarlane PS, Callander R. *Pathology illustrated*. 3rd ed. Edinburgh: Churchill Livingstone; 1991.

17. Lakhani SR, Dilly SA, Finlayson CJ. *Basic pathology: an introduction to the mechanisms of disease*. 3rd ed. London: Arnold; 2003.

18. Stevenson CE. Statistical models for cancer screening. *Stat Methods Med Res* 1995;**4**:18–32.

19. Lincoln T, Weiss GH. A statistical evaluation of recurrent medical examinations. *Oper Res* 1964; **12**:187–205.

20. Knox E. A simulation system for screening procedures. In Mclachlan G, editor. *Future and present indicatives, problems and progress in medical care*. 9th series. Oxford: Nuffield Provincial Hospitals Trust; 1973. pp. 17–55.

21. Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika* 1969;**56**:601–14.

22. Sonnenberg A, Delco F, Inadomi JM. Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Ann Intern Med* 2000;**133**:647–9.

23. Sonnenberg A, Delco F. Cost-effectiveness of a single colonoscopy in screening for colorectal cancer. *Arch Intern Med* 2002;**162**:163–8.

24. Gyrd-Hansen D, Sogaard J, Kronborg O. Colorectal cancer screening: efficiency and effectiveness. *Health Econ* 1998;**7**:9–20.

25. Gyrd-Hansen D. Is it cost effective to introduce screening programmes for colorectal cancer?: illustrating the principles of optimal resource allocation. *Health Policy* 1997;**41**:189–99.

26. Ladabaum U, Chopra CL, Huang G, Scheiman JM, Chernew ME, Fendrick A. Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer: a cost-effectiveness analysis. *Ann Intern Med* 2001;**135**:769–81.

27. Shimbo, T. Cost-effectiveness analysis of strategies for colorectal cancer screening in Japan. *Int J Technol Assess Health Care* 1994;**10**:359–75.

28. Vijan S, Hwang EW, Hofer TP, Hayward RA. Which colon cancer screening test: a comparison of costs, effectiveness, and compliance. *Am J Med* 2001;**111**:593–601.

29. Frazier AL. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA* 2000;**284**:1954–61.

30. Wagner JD, Tunis S, Brown ML, Ching A, Almeida R. The cost effectiveness of colorectal cancer screening in average-risk adults. In Young GP, Rozen R, Levin B, editors. *Prevention and early detection of colorectal cancer*, Philadelphia, PA: Saunders; 1996. pp. 321–56.

31. US Office of Technology Assessment. *The cost-effectiveness of colorectal cancer screening in average-risk adults*. OTA-BP-H-146. Washington, DC: US Office of Technology Assessment; 1995.

32. Neilson AR, Whynes DK. Cost-effectiveness of screening for colorectal cancer: a simulation model. *IMA J Math Appl Med Biol* 1995;**12**:355–67.

33. Khandker RK, Dulski JD, Kilpatrick JB, Ellis RP, Mitchell JB. A decision model and cost-effectiveness analysis of colorectal cancer screening and surveillance guidelines for average-risk adults. *Int J Technol Assess Health Care* 2000;**16**:799–810.

34. Eddy DM. Screening for colorectal cancer. *Ann Intern Med* 1990;**113**:373–84.

35. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN–COLON simulation model for the evaluation of colorectal cancer screening. *Comput Biomed Res* 1999;**32**:13–33.

36. Loeve F. Endoscopic colorectal cancer screening: a cost-saving analysis. *J Nat Cancer Inst* 2000;**92**:557–63.

37. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed* 1985;**20**:79–93.

38. Ness RM, Holmes AM, Klein R, Dittus R. Cost–utility of one-time colonoscopic screening for colorectal cancer at various ages. *Am J Gastroenterol* 2000;**95**:1800–11.

39. Allen MW, Hendi P, Bassett L, Phelps ME, Gambhir SS. A study on the cost effectiveness of sestamibi scintimammography for screening women with dense breasts for breast cancer. *Breast Cancer Res Treat* 1999;**55**:243–58.

40. Mandelblatt J. The costs and effects of cervical and breast cancer screening in a public hospital emergency room. *Am J Public Health* 1997;**87**:1182–9.

41. Lindfors KK, Rosenquist CJ. Needle core biopsy guided with mammography: a study of cost-effectiveness. *Radiology* 1994;**190**:217–22.

42. Lindfors KK, Rosenquist J. The cost-effectiveness of mammographic screening strategies. *JAMA* 1995;**274**:881–4.

43. Salzmann P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. *Ann Intern Med* 1997;**127**:955–65.

44. Lai MS. Efficacy of breast-cancer screening for female relatives of breast-cancer-index cases: Taiwan multicentre cancer screening (TAMCAS). *Int J Cancer* 1998;**78**:21–6.

45. van Oortmarssen GJ, Habbema JD, van der Mass PJ, de Koning H, Collette HJA, Verbeek ALM, *et al*. A model for breast cancer screening. *Cancer* 1990;**66**:1601–12.

46. van Oortmarssen GJ, Habbema JDF, Lubbe J Th N, van der Mass PJ. A model-based analysis of the HIP project for breast cancer screening. *Int J Cancer* 1990;**46**:207–13.

47. Szeto KL, Devlin NJ. The cost-effectiveness of mammography screening: evidence from a microsimulation model for New Zealand. *Health Policy* 1996;**38**:101–15.

48. Eddy DM. Screening for breast cancer. *Ann Intern Med* 1990;**111**:389–99.

49. Schwartz M. An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer* 1978;**41**:1550–64.

50. Schwartz M. A mathematical model used to analyse breast cancer screening strategies. *Oper Res* 1978;**26**:937–55.

51. Parmigiani G. *Decision models for screening in breast cancer*. 2003. URL: http://www.biostat.jhsph.edu/biostat/event/seminar/semparmigiani.shtml. Accessed 30 June 2003.

52. Parmigiani G. *Modeling in medical decision making. A Bayesian approach*. New York: Wiley; 2002.

53. Jenkins D, Sherlaw-Johnson C, Gallivan S. Can papilloma virus testing be used to improve cervical cancer screening? *Int J Cancer* 1996;**65**:768–73.

54. Fahs MC. Cost effectiveness of cervical cancer screening for the elderly. *Ann Intern Med* 1992;**117**:520–7.

55. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. *Am J Epidemiol* 2000;**151**:1158–71.

56. Taylor LA, Sorensen SV, Ray NF, Halpern MT, Harper DM. Cost-effectiveness of the conventional Papanicolaou test with a new adjunct to cytological screening for squamous cell carcinoma of the uterine cervix and its precursors. *Arch Fam Med* 2000;**9**:713–21.

57. Goldie SJ, Weinstein MC, Kuntz KM, Freedberg KA. The costs, clinical benefits, and cost-effectiveness of screening for cervical cancer

in HIV-infected women. *Ann Int Med* 1999;
130:97–107.

58. Maxwell GL, Carlson JW, Ochoa M, Krivak T, Rose GS, Myers ER. Costs and effectiveness of alternative strategies for cervical cancer screening in military beneficiaries. *Obst Gynecol* 2002; **100**:740–8.

59. Eddy DM. Screening for cervical cancer. *Ann Intern Med* 1990;**113**:214–26.

60. Habbema JDF, van Oortmarssen GJ, Lubbe J Th N, van der Mass PJ. Model building on the basis of Dutch cervical cancer screening data. *Maturitas* 1985;**7**:11–20.

61. van den Akker-van Marle ME, van Ballegooÿen M, van Oortmarssen GJ, Boer R, Habbema JDF. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *J Nat Cancer Instit* 2002;**94**:193–204.

62. Parkin DM. A computer simulation model for the practical planning of cervical cancer screening programmes. *Br J Cancer* 1985;**51**:551–68.

63. Gustafsson L, Adami HO. Natural history of cervical neoplasia: consistent results obtained by an identification technique. *Br J Cancer* 1989; **60**:132–41.

64. Gustafsson L, Adami HO. Optimization of cervical cancer screening. *Cancer Causes Control* 1992; **3**:125–36.

65. Barton P, Bryan S. A simulation model to explore the cost-effectiveness of policy changes to the UK cervical screening programme. Paper presented at Health Economists' Study Group, Brunel University, July 2002.

66. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ* 2003;**12**:837–48.

67. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. Impact of systematic false-negative test results on the performance of faecal occult blood screening. *Eur J Cancer* 2001;**37**:912–17.

68. Coley CM, Barry D, Fleming C. Early detection of prostate cancer: Part II: estimating the risks, benefits, and costs. *Ann Intern Med* 1997; **126**:468–79.

69. Church TR. Simulated screening for prostate cancer: the useful model. *J Nat Cancer Inst* 2003; **95**:838–9.

70. Day NE, Walter SD. Simplified models for screening for chronic disease: estimation procedures for mass screening programmes. *Biometrics* 1984;**40**:1–14.

71. Spratt JS, Greenberg RA, Heuser LS. Geometry, growth rates, and duration of cancer and carcinoma in situ of the breast before detection by screening, *Cancer Res* 1986;**46**:970–4.

72. Peer P, van Dijck JAAM., Hendricks J, Holland R, Verbeek ALM. Age-dependent growth rate of primary breast cancer. *Cancer* 1993;**71**:3547–51.

73. Duffy SW, Chen H-H, Prevost TC, Tabar L. Markov chain models of breast cancer progression and its arrest by screening. In Duffy SW, Hill C, Esteve J, editors. *Quantitative methods for the evaluation of cancer screening*. London: Arnold; 2001. pp. 42–60.

74. Duffy SW, Day NE, Tabar L, Chen HH. Markov models of breast tumour progression: some age-specific results, *J Nat Cancer Inst* 1997;**22**:93–7.

75. Duffy SW, Chen HH, Tabar L, Fagerberg G, Paci E. Sojourn time, sensitivity and positive predictive value of mammography screening for breast cancer in women aged 40–49. *Int J Epidemiol* 1996; **25**:1139–45.

76. Prevost TC, Launoy G, Duffy SW, Chen H-H. Estimating sensitivity and sojourn time in screening for colorectal cancer, a comparison of statistical approaches, *AmJ Epidemiol* 1998; **148**:609–19.

77. Madan J. The use of mathematical modelling to evaluate breast cancer screening. *MSc Dissertation*. University of Salford; 2004.

78. Turnbull RB, Kyle K, Watson FR, Spratt J. Cancer of the colon: the influence of the no touch isolation technique on survival rates. *Ann Surg* 1967;**166**:420–7.

79. Dukes CE. The classification of cancer of the rectum. *J Pathol Bacteriol* 1932;**35**:323–32.

80. Rickert RR, Auerbach O, Garfinkel L, Hammond EC, Frasca JM. Adenomatous lesions of the large bowel: an autopsy survey. *Cancer* 1979; **43**:1847–57.

81. Arminski TC, McLean DW. Incidence and distribution of adenomatous polyps of the colon and rectum based on 1,000 autopsy examinations. *Dis Colon Rectum* 1964;**7**:249–61.

82. Blatt LJ. Polyps of the colon and rectum: incidence and distribution. *Dis Colon Rectum* 1961;**4**:277–82.

83. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. *Cancer* 1982;**49**:819–25.

84. Eide TJ, Stalsberg H. Polyps of the large intestine in northern Norway. *Cancer* 1978;**42**:2839–48.

85. Williams AR, Balasooriya BAW, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool. *Gut* 1982;**23**:835–42.

86. Office for National Statistics. *Cancer statistics: registrations of cancer diagnosed in 2000, England*. Series MB1 No. 31. London: Office for National Statistics; 2003.

87. Whynes DK, Walker AR, Hardcastle JD. Cost savings in mass population screening for colorectal cancer: is it cost-effective? *Health Econ* 1992; **1**:53–60.

88. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarthy RL. Natural history of untreated colonic polyps. *Gastroenterology* 1987; **93**:1009–13.

89. Hoff G, Foerster MH, Vatn MH, Sauar J, Larsen S. Epidemiology of the polyps in the rectum and colon. Recovery and evaluation of unresected polyps 2 years after detection. *Scand J Gastroenterol* 1986;**21**:853–62.

90. Hofstad B, Vatn MH, Larsen S, Osnes M. Growth of colorectal polyps: recovery and evaluation of unresected polyps of less than 10 mm, 1 year after detection. *Scand J Gastroenterol* 1994;**24**:640–5.

91. Knoernschild MD. Growth rate and malignant potential of colonic polyps: early results. *Surg Forum* 1963;**14**:137–8.

92. Myles JP, Nixon RM, Duffy SW, Tabar L, Boggis C, Evans G, *et al*. Bayesian evaluation of breast cancer screening using data from two studies. *Stat Med* 2003;**22**:1661–74.

93. Dinning JP, Hixson LJ, Clark LC. Prevalence of distal colonic neoplasia associated with proximal colon cancers. *Arch Intern Med* 1994;**154**:853–6.

94. Levin TR, Palitz A, Grossman S, Conell C, Finkler C, Ackerson L, *et al*. Predicting advanced proximal colonic neoplasia with screening sigmoidoscopy. *JAMA* 1999;**281**:1611–17.

95. Goldie S, Kuntz KM. A potential error in evaluating cancer screening: a comparison of 2 approaches for modelling underlying disease progression. *Med Decis Making* 2003;**23**:232–41.

96. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, *et al*. Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 1993;**329**:1977–81.

97. Gatto NM, Frucht H, Sundararajan V, Jacobson JS, Grann VR, Neugut AI. Risk of perforation after colonoscopy and sigmoidoscopy: a population-based study. *J Nat Cancer Inst* 2003;**95**:230–6.

98. Winawer SJ, Zauber AG, O'Brien MJ, Ho MN, Gottlieb L, Sternberg SS, *et al*. Randomised comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. *N Engl J Med* 1993; **328**:901–6.

99. Blyth CR, Still HA. Binomial confidence intervals. *J Am Stat Assoc* 1983;**78**:108–16.

100. Windeler J, Kobberling J. Colorectal carcinoma and Haemoccult: a study of its value in mass screening using meta-analysis. *Int J Colorect Dis* 1987;**2**:22–8.

101. Allison JE, Feldman R, Tekawa IS. Hemoccult screening in detecting colorectal neoplasm: sensitivity, specificity, and predictive value. *Ann Intern Med* 1990;**112**:328–33.

102. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. *N Engl J Med* 1996; **334**:155–9.

103. UK CRC Screening Pilot Evaluation Team. *Evaluation of the UK colorectal cancer screening pilot*. 2003. http://www.cancerscreening.nhs.uk/bowel/ finalreport.pdf. Accessed 8 November 2007.

104. Atkin WS, Rogers P, Cardwell C, Cook C, Cuzick J, Wardle J, *et al*. Wide variation in adenoma detection rates at screening flexible sigmoidoscopy. *Gastroenterology* 2004;**126**:1247–56.

105. Hixson LJ, Fennerty MB, Sampliner RE, Garewal HS. Prospective blinded trial of the colonoscopic miss-rate of large colorectal polyps. *Gastrointest Endosc* 1991;**37**:125–7.

106. Rex DK, Cutler CS, Lemmel GT, Rahmani EY, Clark DW, Helper DJ, *et al*. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 1997;**112**:24–8.

107. Bressler B, Paszat L, Vinden C, He J, Li C, Rabeneck L. Colonoscopic miss rates for colorectal cancer: a population based analysis. *Gastrointest Endosc* 2004;**59**:AB110.

108. Selby JV, Friedman GD, Quesenberry CP, Weiss NS. A case–control study of sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992;**326**:654–7.

109. Atkin WS. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. *Lancet* 2002; **359**:1291–300.

110. Levin TR, Conell C, Shapiro JA, Chazan SG, Nadel MR, Selby JV. Complications of screening flexible sigmoidoscopy. *Gastroenterology* 2002; **123**:1786–92.

111. Government Actuary's Department. *Interim life tables*. 2004. URL: http: www.gad.gov.uk/. Accessed 4 February 2004.

112. South West Cancer Intelligence Service. *Audit study of patients with colorectal cancer undertaken in the Wessex Region 1991–1995*. Bristol: South West Cancer Intelligence Service; 1995.

113. Hardcastle JD. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;**348**:1472–7.

114. Lund JN, Scholefield JH, Grainge MJ, Smith SJ, Mangham C, Armitage NC, *et al*. Risks, costs, and compliance limit colorectal adenoma surveillance: lessons from a randomised trial. *Gut* 2001; **49**:91–6.

115. Sanderson H, Walker A, Young D. Colorectal cancer. In Stevens A, Raftery J, Mant J, Simpson S, editors. *Health care needs assessment*, First Series, 2nd ed., Vol. 1. Abingdon: Radcliffe Publishing; 2004. pp. 449–502.

116. Department of Health. *NHS reference costs*. London: Department of Health; 2003.

117. Netten A, Curtis L. *Unit costs of health and social care*. Canterbury: PSSRU; 2004.

118. Whynes DK, Frew E, Edwards R, Atkin WS. Costs of flexible sigmoidoscopy screening for colorectal cancer in the United Kingdom. *Int J Technol Assess Health Care* 2003;**19**:384–95.

119. Lloyd Jones ML, Hummel S, Bansback N, Orr B, Seymour M. A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer. *Health Technol Assess* 2001;**5**(25).

120. Saltz LB, Cox JV, Rosen LS, Fehrenbacher L, Moore MJ, Maroun JA, *et al*. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *New Eng J Med* 2000;**343**:905–14.

121. National Audit Office. The performance of the NHS cervical screening programme in England: a report by the Comptroller and Auditor General. London: National Audit Office; 1998.

122. UK Colorectal Cancer Screening Pilot Group Results of the first round of a demonstration pilot of screening for colorectal cancer in the United Kingdom. *BMJ* 2004;**329**:133

123. Whynes DK, Neilson AR. Convergent validity of two measures of the quality of life. *Health Economics* 1993;**2**:229–35.

124. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. *American Journal of Gastroenterology* 1999; **94**:1650–7.

125. Dominitz JA, Provenzale D. Patient preferences and quality of life associated with colorectal cancer screening. *Am J Gastroenterol* 1997;**92**:2171–8.

126. Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RAM, Schroder FH, *et al*. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European randomised study of screening for prostate cancer. *J Nat Cancer Inst* 2003;**95**:868–78.

127. Bazaraa MS, Sherali HD, Shetty HD. *Nonlinear programming: theory and algorithms*. New York: Wiley; 1993.

128. Church TR. Simulated screening for prostate cancer: the useful model. *J Nat Cancer Inst* 2003; **95**:838–9.

129. Office for National Statistics. *Mortality statistics: review of the Registrar General on deaths by cause, sex and age in England and Wales, 2002*. Series DH2 No. 29. London: Office for National Statistics; 2003.

130. Tappenden P, Eggington S, Nixon R, Chilcott J, Sakai H, Karnon J. *Colorectal cancer screening options appraisal. Cost-effectiveness, cost–utility and resource impact of alternative screening options for colorectal cancer*. Report to the English Bowel Cancer Screening Working Group. Sheffield: ScHARR; 2004.

131. Atkin WS, Saunders BP. Surveillance guidelines after removal of colorectal adenomatous polyps. *Gut* 2002;**51**:v6–9.

132. Whynes DK, Frew EJ, Manghan CM, Scholefield JH, Hardcastle JD. Colorectal cancer, screening and survival: the influence of socio-economic deprivation. *Public Health* 2003;**117**:389–95.

133. Karnon J, Brown J. Tamoxifen plus chemotherapy versus tamoxifen alone as adjuvant therapies for node-positive postmenopausal women with early breast cancer: a stochastic economic evaluation. *PharmacoEconomics* 2002;**20**:119–37.

134. Hillner BE. Benefit and projected cost-effectiveness of anastrozole versus tamoxifen as initial adjuvant therapy for patients with early-stage estrogen receptor-positive breast cancer. *Cancer* 2004;**101**:1311–22.

135. Karnon J, Jones T. A stochastic economic evaluation of letrozole versus tamoxifen as a first-line hormonal therapy for advanced breast cancer in postmenopausal patients. *PharmacoEconomics* 2003;**21**:513–25.

136. Hutton J, Brown R, Borowitz M, Abrams K, Rothman M, Shakespeare A. A new decision model for cost–utility comparisons of chemotherapy in recurrent metastatic breast cancer. *PharmacoEconomics* 1996;**9**:8–22.

137. Scholefield JH, Moss S, Sufi F, Mangham C, Hardcastle JD. Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomised controlled trial. *Gut* 2002;**50**:840–4.

138. Karnon J, Brown J, Zeuner D, Briggs A. *Issues in the economic evaluation of antenatal screening programmes for genetic disorders*. HERG Discussion Paper No. 18. Brunel University; 1997.

139. Segal L, Dalton AC, Richardson J. Cost-effectiveness of primary prevention of non-insulin dependent diabetes mellitus. *Heatlh Promot Int* 1998;**13**:197–209.

140. CDC Diabetes Cost-Effectiveness Study Group. The cost-effectiveness of screening for type 2 diabetes. *JAMA* 1998;**280**:1757–63.

141. Chen THH, Yen MF, Tung TH. A computer simulation model for cost-effectiveness analysis of mass screening for type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2001;**54** (Suppl 1): S37–42.

131

142. Raikou M, McGuire A. The economics of screening and treatment of type 2 diabetes mellitus. *PharmacoEconomics* 2003;**21**:543–64.

143. Javitt JC, Canner JK, Sommer A. Cost effectiveness of current approaches to the control of retinopathy in type I diabetics. *Ophthalmology* 1989;**96**:225–64.

144. Javitt JC, Canner JK, Frank RG, Steinwachs GM, Sommer A. Detecting and treating retinopathy in patients with type I diabetes mellitus. A health policy model. *Ophthalmology* 1990;**97**:483–95.

145. Dasbach EJ, Fryback DG, Newcomb PA, Klein R, Klein BE. Cost effectiveness of strategies for detecting diabetic retinopathy. *Med Care* 1991;**29**:20–39.

146. Javitt JC, Aiello LP, Bassi LJ, Chiang YP, Canner JK. Detecting and treating retinopathy in patients with type I diabetes mellitus. Savings associated with improved implementation of current guidelines. American Academy of Ophthalmology. *Ophthalmology* 1991;**98**:1565–74.

147. Fendrick A, Javitt JC, Chiang YP. Cost-effectiveness of the screening and treatment of diabetic retinopathy. *Int J Technol Assess Health Care* 1992;**8**:694–707.

148. Javitt JC, Aiello LP, Chiang YP, Ferris FL, Canner JK, Greenfield S. Preventive eye care in people with diabetes is cost saving to the federal government; implications for health care reform. *Diabetes Care* 1994;**17**:909–17.

149. Javitt JC, Aiello LP. Cost-effectiveness of detecting and treating diabetic retinopathy. *Ann Intern Med* 1996;**124**:164–9.

150. Mellanby A, Milne R. *Reducing the interval for diabetic retinal screening*. Development and Evaluation Committee Report No. 107. Southampton: Wessex Institute for Health Research and Development; 1999.

151. Vijan S, Hofer TP, Hayward RA. Cost–utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA* 2000;**283**:889–96.

152. Crijns H. Continuous computer simulation analysis of the cost-effectiveness of screening and treating diabetic retinopathy. *Int J Technol Assess Health Care* 1999;**15**:198–206.

153. Davies R, Brailsford S, Roderick P, Canning C, Crabbe D. Using simulation modelling for evaluating screening services for diabetic retinopathy. *J Oper Res Soc* 2000;**51**:476–84.

154. James M, Little R. *Diabetic retinopathy. Report to the National Screening Committee*. Staffordshire: Keele University; 2001.

155. Brailsford SC, Davies R, Canning C, Roderick PJ. Evaluating screening policies for the early detection of retinopathy in patients with non-

156. Faronato P, de Bigontina G. A cost–benefit analysis of two mass screening strategies for albuminuria in diabetic patients. *Diabetes Nutri Metab* 1994;**7**:325–9.

157. Borch-Johnsen K. Is screening and intervention for microalbuminuria worthwhile in patients with insulin dependent diabetes? *BMJ* 1993;**306**:1722–3.

158. Kiberd BA, Jindal KK. Screening to prevent renal failure in insulin-dependent diabetic patients; an economic evaluation. *BMJ* 1995;**311**:1595–9.

159. Finkelstein EA, Troped PJ, Will JC, Palombo R. Cost-effectiveness of a cardiovascular disease risk reduction program aimed at financially vulnerable women: the Massachusetts WISEWOMAN project. *J Womens Health Gender-Based Med* 2002;**11**:519–26.

160. Marks D. Cost-effectiveness analysis of different approaches of screening for familial hypercholesterolaemia. *BMJ* 2002;**324**:1303–6.

161. Marshall T, Rouse A. Resource implications and health benefits of primary prevention strategies for cardiovascular disease in people aged 30 to 74: mathematical modelling study. *BMJ* 2002;**325**:197–202.

162. Moskowitz H, Plante R, Tsai HT. A multistage screening model for evaluation and control of misclassification error in the detection of hypertension. *Manage Sci* 1993;**39**:307–21.

163. Junod AF. Should there be systematic screening of coronary heart disease in asymptomatic patients with risk factors alone? A decision analysis approach. *Diabetes Metab* 1998;**24**:496–507.

164. Sox HC. The role of exercise testing in screening for coronary artery disease. *Ann Intern Med* 1989;**110**:456–69.

165. Lee TL, Solomon NA, Heidenreich PA, Oehlert J, Garber AM. Cost-effectiveness of screening for carotid stenosis in asymptomatic persons. *Ann Intern Med* 1997;**126**:337–46.

166. Derdeyn CP, Powers WJ. Cost-effectiveness of screening for asymptomatic carotid atherosclerotic disease. *Stroke* 1996;**27**:1944–50.

167. Yin DP, Carpenter JP. Cost-effectiveness of screening for asymptomatic carotid stenosis. *J Vasc Surg* 1998;**27**:245–55.

168. Yoshimoto Y, Wakai S. Cost-effectiveness analysis of screening for asymptomatic, unruptured intracranial aneurysms: a mathematical model. *Stroke* 1999;**30**:1621–7.

169. Baba Y, Takahashi M, Korogi Y. Decision analysis of cost-effectiveness of magnetic resonance angiography for mass screening for intracranial aneurysms. *Acad Radiol* 1998;**5** (Suppl 2):S297–9.

insulin dependent diabetes. *Health Care Manage Sci* 1998;**1**:115–24.

170. Frame PS, Fryback DG, Patterson C. Screening for abdominal aortic aneurysm in men aged 60 to 80 years: a cost-effectiveness analysis. *Ann Intern Med* 1993;**119**:411–16.

171. Mason JM. Population screening for abdominal aortic aneurysm: do the benefits outweigh the costs. *J Public Health Med* 1993;**15**:154–60.

172. Pentikainen TJ, Sipila T, Rissanen P, Soisalon-Soininen S, Salo J. Cost-effectiveness of targeted screening for abdominal aortic aneurysm: Monte Carlo-based estimates. *Int J Technol Assess Health Care* 2000;**16**:22–34.

173. Soisalon-Soininen S, Rissanen P, Pentikainen TJ, Salo JA. Cost-effectiveness of screening for familial abdominal aortic aneurysms. *Vasa* 2001;**30**:262–70.

174. St Leger AS, Spencely M, McCollum CN, Mossa M. Screening for abdominal aortic aneurysm: a computer assisted cost-utility analysis. *Eur J Vasc Endovasc Surg* 1996;**11**:183–90.

175. Last JM. *A dictionary of epidemiology.* 4th ed. New York: Oxford University Press; 2001.

176. Sarasin FP, Bounameaux H. Antithrombotic strategy after total hip replacement: a cost-effectiveness analysis comparing prolonged oral anticoagulants with screening for deep vein thrombosis. *Arch Intern Med* 1996;**156**:1661–8.

177. Derdeyn CP, Gage BF, Grubb RL, Powers WJ. Cost-effectiveness analysis of therapy for symptomatic carotid occlusion: PET screening before selective extracranial-to-intracranial bypass versus medical treatment. *J Nucl Med* 2000; **41**:800–7.

178. Glance LG. Selective preoperative cardiac screening improves five-year survival in patients undergoing major vascular surgery: a cost-effectiveness analysis. *J Cardiothorac Vasc Anesth* 1999;**13**:265–71.

179. Butler WE, Barker FG, Crowell RM. Patients with polycystic kidney disease would benefit from routine magnetic resonance angiographic screening for intracerebral aneurysms:a decision analysis. *Neurosurgery* 1996;**38**:506–15.

180. Jespersen HF, Jorgensen HS, Nakayama H, Reith J, Olsen TS. Carotid Doppler – costs and need after stroke or TIA. *Acta Neurol Scand* 2002; **105**:1–4.

181. Law MR, Morris J, Wald NJ. Screening for abdominal aortic aneurysms. *J Med Screen* 1994; **1**:110–15.

182. Lee TY, Korn P, Heller JA, Kilaru S, Beavers FP, Bush HL, *et al.* The cost-effectiveness of a "quick-screen" program for abdominal aortic aneurysms. *Surgery* 2002;**132**:399–407.

183. Lederle FA, Johnson GR, Wilson SE, Littooy FN, Krupski WC, Bandyk D, *et al.* Yield of repeated screening for abdominal aortic aneurysm after a 4-year interval. Aneurysm Detection and Management Veterans Affairs Cooperative Study Investigators. *Arch Intern Med* 2000;**160**:1117–21.

184. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *PharmacoEconomics* 1998;**13**:397–409.

185. Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RA, *et al.* The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *Lancet* 2002;**360**:1531–9.

186. Cronenwett JL, Sargent SK, Wall MH, Hawkes ML, Freeman DH, Dain BJ, *et al.* Variables that affect the expansion rate and outcome of small abdominal aortic aneurysms. *J Vasc Surg* 1990; **11**:260–8.

187. Chichester Aneurysm Screening Group, Viborg Aneurysm Screening Study, Western Australian Abdominal Aortic Aneurysm Program, Multicentre Aneurysm Screening Study. A comparative study of the prevalence of abdominal aortic aneurysms in the United Kingdom, Denmark, and Australia. *J Med Screen* 2001;**8**:46–50.

188. Sin K, Bonaa KH, Jacobsen BK, Bjork L, Solberg S. Prevalence of and risk factors for abdominal aortic aneurysms in a population-based study: the Tromso Study. *Am J Epidemiol* 2001;**154**:236–44.

189. Lederle FA, Johnson GR, Wilson SE, Chute EP, Littooy FN, Bandyk D, *et al.* Prevalence and associations of abdominal aortic aneurysm detected through screening. Aneurysm Detection and Management (ADAM) Veterans Affairs Cooperative Study Group. *Ann Intern Med* 1997; **126**:441–9.

190. Vardulaki KA, Prevost TC, Walker NM, Day NE, Wilmink AB, Quick CR, *et al.* Growth rates and risk of rupture of abdominal aortic aneurysms. *Br J Surg* 1998;**85**:1674–80.

191. Englund R, Hudson P, Hanel K, Stanton A. Expansion rates of small abdominal aortic aneurysms. *Aust N Z J Surg* 1998;**68**:21–4.

192. Cook TA, Galland RB. A prospective study to define the optimum rescreening interval for small abdominal aortic aneurysm. *Cardiovasc Surg* 1996; **4**:441–4.

193. Stonebridge PA, Draper T, Kelman J, Howlett J, Allan PL, Prescott R, *et al.* Growth rate of infrarenal aortic aneurysms. *Eur J Vasc Endovasc Surg* 1996;**11**:70–3.

194. Michaels JA. Use of mortality rate after aortic surgery as a performance indicator. *Br J Surg* 2003;**90**:827–31.

195. Multicentre Aneurysm Screening Study Group. Multicentre aneurysm screening study (MASS): cost effectiveness analysis of screening for abdominal aortic aneurysms based on four year

results from randomised controlled trial. *BMJ* 2002;**325**:1135–41.

196. Briggs AH. Handling uncertainty in cost-effectiveness models. *PharmacoEconomics* 2000; **17**:479–500.

197. Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN. Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis. *Health Technol Assess* 1999; **3**(11).

198. Cronin EK, Normand C, Henthorn JS, Graham V, Davies SC. Organisation and cost-effectiveness of antenatal haemoglobinopathy screening and follow up in a community-based programme. *Br J Obstet Gynaecol* 2000;**107**:486–91.

199. Phelan L. An analysis of relative costs and potential benefits of different policies for antenatal screening for B thalassaemia trait and varient haemoglobins. *J Clin Pathol* 1999;**52**:697–700.

200. Gallivan S, Jit M, Utley M, Lilford RJ. *A stochastic model to evaluate options for antenatal genetic screening*. CORU Working Paper – CORU Research Report 672. London: University College; 2003.

201. Gill M. An economic appraisal of screening for Down's syndrome in pregnancy using maternal age and serum alpha fetoprotein concentration. *Soc Sci Med* 1987;**24**:725–31.

202. Shackley P. An economic appraisal of alternative pre-natal screening programmes for Down's syndrome. *J Public Health Med* 1993;**15**:175–84.

203. Ganiats TG, Halverson AL, Bogart MH. Incremental cost-effectiveness of incorporating oestriol evaluation in Down syndrome screening programmes. *Prenat Diagn* 1994;**14**:527–35.

204. Fletcher J. Using decision analysis to compare policies for antenatal screening for Down's syndrome. *BMJ* 1995;**311**:351–6.

205. Wald NJ, Kennard A, Hackshaw A, McGuire A. Antenatal screening for Down's syndrome: a review. *Health Technol Assess* 1998;**2**(1).

206. Vintzileos AM, Ananth CV, Fisher AJ, Smulian JC, Day-Salvatore D, Beazoglou T. An economic evaluation of first-trimester genetic sonography for prenatal detection of Down syndrome. *Obstet Gynaecol* 1998;**91**:535–9.

207. Waitzman NJ, Romano PS, Scheffler RM, Harris JA. Estimates of the economic costs of birth defects and cerebral palsy – United States. *Morb Mortal Wkly Rep* 1995;**44**:694–9.

208. Beazoglou T, Heffley D, Kyriopoulos J, Vintzileos A, Benn P. Economic evaluation of prenatal screening for Down syndrome in the USA. *Prenat Diagn* 1998;**18**:1241–52.

209. Benn PA, Rodis JF, Beazoglou T. Cost-effectiveness of estimating gestational age by ultrasonography in Down syndrome screening. *Obstet Gynecol* 1999; **94**:29–33.

210. Seror V, Costet N. Down syndrome serum marker screening: decision criteria and implicit values. *Health Policy* 1998;**43**:83–96.

211. Christiansen M, Larsen SO. An increase in cost-effectiveness of first trimester maternal screening programmes for fetal chromosome anomalies is obtained by contingent testing. *Prenat Diagn* 2002; **22**:482–6.

212. Gallivan S, Utley M. Evaluation of strategies for Downs syndrome screening. *J Oper Res Soc* 2000; **51**:272–7.

213. Ecker JL. The cost-effectiveness of human immunodeficiency virus screening in pregnancy. *Am J Obstet Gynecol* 1996;**174**:716–21.

214. Nakchbandi IA, Longenecker C, Latta RA, Healton C, Smith DG. A decision analysis of mandatory compared with voluntary HIV testing in pregnant women. *Ann Intern Med* 1998; **128**:760–7.

215. Myers ER, Thompson JW, Simpson K. Cost-effectiveness of mandatory compared with voluntary screening for human immunodeficiency virus in pregnancy. *Obstet Gynecol* 1998;**91**:174–81.

216. Ades AE, Sculpher MJ, Gibb DM, Gupta R, Ratcliffe J. Cost effectiveness analysis of antenatal HIV screening in United Kingdom. *BMJ* 1999; **319**:1230–4.

217. Briggs A, Gibb D, Sculpher M, MacDonagh S, Ades T. *Antenatal testing for HIV infection. Developing a model for economic evaluation*. HERG Discussion Paper No. 14. Brunel University; 1996.

218. Postma MJ, Beck EJ, Mandalia S, Sherr L, Walters MDS, Houweling H, *et al*. Universal HIV screening of pregnant women in England: cost effectiveness analysis. *BMJ* 1999;**318**:1656–60.

219. Zaric GS, Bayoumi AM, Brandeau ML, Owens DK. The cost effectiveness of voluntary prenatal and routine newborn HIV screening in the United States. *J Acquir Immune Defic Syndr* 2000;**25**:403–16.

220. Immergluck LC, Cull WL, Schwartz A, Elstein AS. Cost-effectiveness of universal compared with voluntary screening for human immunodeficiency virus among pregnant women in Chicago. *Pediatrics* 2000;**105**:E541–9.

221. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manage Sci* 1998;**1**:133–40.

222. Johnston K, Brown J, Gerard K, O'Hanlon M, Morton A. Valuing temporary and chronic health states associated with breast screening. *Soc Sci Med* 1998;**47**:213–22.

223. Cairns J, Shackley P. Sometimes sensitive, seldom specific: a review of the economics of screening. *Health Econ* 1993;**2**:43–53.

224. Lange M, Norby S, Mooney G. *Developing a cost benefit approach to prenatal analysis for ADPKD*. Health Economics Research Unit Discussion Paper. University of Aberdeen; 1990.

225. Mooney G, Lange M. *Economic appraisal in pre-natal screening: some methodological issues*. Health Economics Research Unit Discussion Paper. University of Aberdeen; 1990.

226. Harvei S. Epidemiology of prostatic cancer. *Tidsskr Nor Laegeforen* 1999;**119**:3589–94.

227. Etzioni R, Cha R, Feuer EJ, Davidov O. Asymptomatic incidence and duration of prostate cancer. *Am J Epidemiol* 1998;**148**:775–85.

228. *Harvard catalogue of preference scores*. URL: http://www.hsph.harvard.edu/cearegistry/. Accessed 16 February 2006.

229. Speight J. Assessing the impact of diabetes screening on quality of life or quality of health?: semantics are important. *Diabetes Care* 2002; **23**:1893–4.

230. Whynes DK, Frew E, Wolstenholme JL. A comparison of two methods for eliciting contingent valuations of colorectal cancer screening. *J Health Econ* 2003;**22**:555–74.

231. Kimball HR. Screening for disease in asymptomatic patients. *Trans Am Clin Climatol Assoc* 1987;**99**:23–31.

232. Barratt AL, Irwig LM, Glasziou PP, Salkeld GP, Houssan N. Benefits, harms and costs of screening mammography in women 70 years and over: a systematic review. *Med J Aust* 2002;**176**:266–72.

233. Cairns J, Shackley P. *Tries hard but could do better: a report on the economics of screening*. Health Economics Research Unit Discussion Paper. University of Aberdeen; 1992.

234. Hall J. Best medical practice in practice: measuring efficiency in mammography screening. *Int J Health Planning Manage* 1989;**4**:235–46.

235. Johnson WO, Gastwirth JL, Perason LM. Screening without a 'gold standard': the Hui–Walter paradigm revisited. *Am J Epidemiol* 2001;**153**:921–4.

236. Shaheen NJ, Provenzale D. Screening strategies in gastroesophageal reflux disease: early identification of esophageal carcinoma. *Adv Intern Med* 2001;**47**:137–57.

237. Boll APM, Severens JL, Verbeek ALM, van der Vliet JA. Mass screening on abdominal aortic aneurysm in men aged 60 to 65 years in The Netherlands. Impact on life expectancy and cost-effectiveness using a Markov model. *Eur J Vasc Endovasc Surg* 2003;**26**:74–80.

238. Birkmeyer JD, AuBuchon JP, Littenberg B, O'Connor GT, Nease RF, Nugent WC, *et al*. Cost-effectiveness of preoperative autologous donation in coronary artery bypass grafting. *Ann Thorac Surg* 1994;**57**:161–9.

239. Weinstein MC, Tosteson ANA. Cost-effectiveness of hormone replacement therapy. *Ann N Y Acad Sci* 1990;**S92**:162–72.

240. Weinstein MC. Estrogen use in postmenopausal women – costs, risks, and benefits. *N Engl J Med* 1980;**303**:308–16.

241. Huber TS, McGorray SP, Carlton LC, Irwin PB, Flug RR, Flynn TC, *et al*. Intraoperative autologous transfusion during elective infrarenal aortic reconstruction: a decision analysis model. *J Vasc Surg* 1997;**25**:984–94.

242. Birkmeyer JD, Goodnough LT, AuBuchon JP, Noordsij PG, Littenberg B. The cost-effectiveness of preoperative autologous blood donation for total hip and knee replacement. *Transfusion* 1993; **33**:544–51.

243. Messori A, Becagli P, Berto V, Trippoli S, Font M, Tosolini S, *et al*. Advanced HIV infection treated with zidovudine monotherapy: lifetime values of absolute cost-effectiveness as a pharmacoeconomic reference for future studies evaluating antiretroviral combination treatments. *Ann Pharmacother* 1997;**31**:1447–54.

244. van der Meulen JHP, Mol BWJ, Pajkrt E, van Lith JMM, Voorn W. Use of the disutility ratio in prenatal screening for Down's syndrome. *Br J Obstet Gynaecol* 1999;**160**:108–15.

245. Boer R, de Koning HJ, van Oortmarssen GJ, van der Mass PJ. In search of the best upper age limit for breast cancer screening. *Eur J Cancer* 1995;**31A**:2040–3.

246. Goyder E, Irwig LM. Screening for type 2 diabetes mellitus: a decision analytic approach. *Diabet Med* 2000;**17**:469–77.

247. Ebell MH, Warbasse L, Brenner C. Evaluation of the dyspeptic patient: a cost–utility study. *J Fam Pract* 1997;**44**:545–55.

248. Krahn M, Naylor CD, Basinski AS, Detsky AS. Comparison of an aggressive (US) and a less aggressive (Canadian) policy for cholesterol screening and treatment. *Ann Intern Med* 1991; **115**:248–55.

249. Adams PC, Gregor JC, Kertesz AE, Valberg LS. Screening blood donors for hereditary hemochromatosis: decision analysis model based on a 30-year database. *Gastroenterology* 1995; **109**:177–88.

250. Adams PC, Kertesz AE, Valberg LS. Screening for hemochromatosis in children of homozygotes: prevalence and cost-effectiveness. *Hepatology* 1995; **22**:1720–7.

251. Littenberg B, Garber AM, Sox HC. Screening for hypertension. *Ann Intern Med* 1990;**112**:192–202.

252. Tosteson ANA, Rosenthal DI, Melton LJ, Weinstein MC. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Ann Intern Med* 1990;**113**:594–603.

253. Schechter CB, Rose DN, Fahs MC, Silver AL. Tuberculin screening: cost-effectiveness analysis of various testing schedules. *Am J Prev Med* 1990; **6**:167–75.

254. Velanovich V. Immediate biopsy versus observation for abnormal findings on mammograms: an analysis of potential outcomes and costs. *Am J Surg* 1995;**170**:327–32.

255. Raab SS, Hornberger J. The effect of a patient's risk-taking attitude on the cost effectiveness of testing strategies in the evaluation of pulmonary lesions. *Chest* 1997;**111**:1583–90.

256. Chen TH-H, Yen M-F, Tung T-H. A computer simulation model for cost-effectiveness analysis of mass screening for type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2001;**54**:S37–42.

257. Singer ME, Younossi ZM. Cost effectiveness of screening for hepatitis C virus in asymptomatic, average-risk adults. *Am J Med* 2001;**111**:614–21.

258. Anderson JP, Moser RJ. Parasite screening and treatment among Indochinese refugees: cost–benefit/utility and the general health policy model. *JAMA* 1985;**253**:2229–35.

259. Romano PS, Waitzman NJ. Can decision analysis help us decide whether ultrasound screening for fetal anomalies is worth it? *Ann N Y Acad Sci* 1998; **847**:154–72.

260. AuBuchon JP, Birkmeyer JD, Busch MP. Cost-effectiveness of expanded human immunodeficiency virus-testing protocols for donated blood. *Transfusion* 1997;**37**:45–51.

261. Åsberg A, Tretli S, Hveem K, Bjeve KS. Benefit of population-based screening for phenotypic hemochromatosis in young men. *Scand J Gastroenterol* 2002;**37**:1212–19.

262. Obuchowski NA, Modic MT, Magdinec MRN, Masaryk TJ. Assessment of the efficacy of non-invasive screening for patients with asymptomatic neck bruits. *Stroke* 1997;**28**:1330–9.

263. Mahadevia PJ, Fleisher LA, Frick KD, Eng J, Goodman SN, Powe NR. Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. *JAMA* 2003;**289**:313–22.

264. Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. *JAMA* 1996;**276**:285–330.

265. Unic I, Verhoef LCG, Stalmeier PFM, van Daal WAJ. Prophylactic mastectomy or screening in women suspected to have the BRCA1/2 mutation: a prospective pilot study of women's treatment choices and medical and decision-analytic recommendations. *Med Decis Making* 2000; **20**:251–62.

266. de Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JCJM, Collette HJA, Hendricks JHCL, *et al*. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer* 1991;**49**:531–7.

267. de Haes JCJM, de Koning HJ, van Oortmarssen GJ, van Agt HME, de Bruyn AE, van der Mass PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer* 1991; **49**:538–44.

268. Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer. A decision analytic view. *JAMA* 1994;**272**:773–80.

269. Shackley P, Cairns J. Evaluating the benefits of antenatal screening: an alternative approach. *Health Policy* 1996;**36**:103–15.

270. Essink-Bot M-L, de Koning HJ, Nijs HGT, Kirkels WJ, van der Maas PJ, Schroder FH. Short-term effects of population-based screening for prostate cancer on health-related quality of life. *J Nat Cancer Inst* 1998;**90**:925–31.

271. Hensley ML, Robson ME, Kauff ND, Korytowsky B, Castiel M, Ostroff J, *et al*. Pre- and postmenopausal high-risk women undergoing screening for ovarian cancer: anxiety, risk perceptions, and quality of life. *Gynecol Oncol* 2003;**89**:440–6.

272. Fry A, Busby-Earle C, Rush R, Cull A. Prophylactic oophprectomy versus screening: psychosocial outcomes in women at increased risk of ovarian cancer. *Psycho-Oncology* 2001;**10**:231–41.

273. Rumbold AR, Crowther CA. Women's experiences of being screened for gestational diabetes mellitus. *Aust N Z J Obstet Gynaecol* 2002;**42**:131–7.

274. Cantor SB, Spann SJ, Volk RJ, Cardenas MP, Warren MM. Prostate cancer screening: a decision analysis. *J Fam Pract* 1995;**41**:33–41.

275. Volk RJ, Cantor SB, Spann SJ, Cass AR, Cardenas MP, Warren MM. Preferences of husbands and wives for prostate cancer screening. *Arch Fam Med* 1997;**6**:72–6.

276. Cormier L, Guillemin F, Valeri A, Fournier G, Cussenot O, Mangin P, *et al*. Impact of prostate cancer screening on health-related quality of life in at-risk families. *Urology* 2002;**59**:901–6.

277. Edelman D, Olsen MK, Dudley TK, Harris AC, Oddone EZ. Impact of diabetes screening on quality of life. *Diabetes Care* 2002;**25**:1022–6.

278. Cairns J, Shackley P, Hundley V. Decision making with respect to diagnostic testing: a method of

valuing the benefits of antenatal screening. *Med Decis Making* 1996;**16**:161–8.

279. Gerard K, Johnston K, Brown J. The role of a pre-scored multi-attribute health classification measure in validating condition-specific health state descriptions. *Health Econ* 1999;**8**:685–99.

280. Dominitz JA, Provenzale D. Patient preferences and quality of life associated with colorectal cancer screening. *Am J Gastroenterol* 1997;**92**:2171–8.

281. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).

282. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992;**31**:301–6.

283. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1987;**150**:782–6.

284. Elstein AS, Chapman GB, Chmiel JS, Knight SJ, Chan C, Nadler RB, *et al*. Agreement between prostate cancer patients and their clinicians about utilities and attribute importance. *Health Expect* 2004;**7**:115–25.

285. Koelemay MJW, Nederkoorn PJ, Reitsma JB, Majoie CB. Systematic review of computed tomographic angiography for assessment of carotid artery disease, *Stroke* 2004;**35**:2306–12.

Appendix 1
Screening modelling search strategies

Searches were conducted in May–June 2003.

Search strategy (example from MEDLINE)

1. exp Technology Assessment, Biomedical/
2. technology assessment.tw.
3. exp Mass Screening/
4. screening.tw.
5. industrial maintenance.tw.
6. 1 or 2 or 3 or 4 or 5
7. exp Cost-Benefit Analysis/
8. cost utilit$.tw.
9. cost effective$.tw.
10. cost benefit$.tw.
11. exp Program Evaluation/
12. exp Research Design/
13. research priorit$.tw.
14. priorit$ research.tw.
15. cost$.ti.
16. economic$.ti.
17. exp Models, Theoretical/
18. exp Models, Econometric/
19. exp Models, Economic/
20. exp Models, Statistical/
21. exp Logistic Models/
22. exp Decision Making/
23. exp Decision Making, Organizational/
24. exp Decision Support Techniques/
25. exp Decision Trees/
26. decision model$.tw.
27. decision analy$.tw.
28. delay time model$.tw.
29. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30. 6 and 29
31. exp Health Services Research/
32. 30 and 31

Sources searched

Electronic databases
- Embase (WebSPIRS) 1980–2003
- IAOR (International Abstracts in Operations Research) (1989–2003)
- INFORMS (Institute for Operations Research and the Management Sciences) Annual Comprehensive Index (ACI) Bibliographic Database (1976–2001)
- INFORMS (Institute for Operations Research and the Management Sciences) Conference Presentations Database (1990–2003)
- OHE (Office of Health Economics) HEED (Health Economic Evaluations Database) (CD-Rom) 1967–2003
- MEDLINE (Ovid) 1966–2003
- NHS EED (Economic Evaluation Database) (The Cochrane Library, 2003, Issue 2).

Other sources
The Internet and the ScHARR Library Catalogue were searched in order to identify discussion papers and other grey literature.

Screening and utility values literature search

Search terms in Ovid MEDLINE from 1966 to February Week 4 2004
1. value of life/.4231
2. quality adjusted life year/.1762
3. quality adjusted life year.tw..1206
4. (qaly$ or qald$ or qale$ or qtime$).tw..919
5. disability adjusted life.tw..190
6. daly$.tw..257
7. health status indicators/.7965
8. (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw..3277
9. (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.577
10. (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw..357
11. sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw..22
12. sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw..241
13. (euroqol or euro qol or eq5d or eq 5d).tw..422
14. (hql or hqol or h qol or hrqol or hr qol).tw..949
15. (hye or hyes).tw..43

16. health$ year$ equivalent$.tw..30
17. health util$.tw..361
18. (hui or hui1 or hui2 or hui3).tw..258
19. disutili$.tw..42
20. rosser.tw..45
21. quality of wellbeing.tw..2
22. qwb.tw..88
23. standard gamble$.tw..286
24. (time trade of$ or time trade-of$).tw..248
25. TTO.tw..153
26. health state util$.tw..53

27. health state val$.tw..59
28. health val$.tw..261
29. visual analog$ scale.tw..6548
30. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29.26101
31. *Mass screening/.21965
32. screening.ti..47612
33. 31 or 32.54391
34. 30 and 33.350

Appendix 2

Summary tables of modelling studies of screening for AAA

Tables 34 and *35* summarise the cost parameters, and the decision parameter values, structural assumptions and input parameters, of the studies reviewed in Chapter 6.

TABLE 34 *Main cost estimates reported by AAA screening studies*

Cost event	Mason, 1993[171]	Frame et *al.*, 1993[170]	Law et *al.*, 1994[181]	St Leger et *al.*, 1996[174]	Soisalon-Soininen et *al.*, 2001[173a]	Lee et *al.*, 2002[182]	MASS[195]
	(UK£)	(US$)	(UK£)	(UK£)	(UK£)	(US$)	(UK£)
Initial screen	?	150	4.6	3.36	105	259 (40)[b]	20.8
Elective surgery	3,000	27,000	4,000	2,371	6,266	16,013	6,909
Emergency surgery	3,600	52,000	6,000	3,914	9,294	28,338	11,176

[a] Cost values are not presented by Pentikainen and colleagues,[172] but may be assumed to be similar to those presented by Soisalon-Soininen and colleagues.[173] The presented costs are converted to UK£ at 8.73 Finnish Marks = £1.
[b] $40 is the cost of the 'quick-screen' evaluated by Lee and colleagues.[182]

TABLE 35 *Screening for abdominal aortic aneurysms modelling study characteristics*

Study	Comparators	Model type	Threshold for surgery	Aneurysm size	Test characteristics	Prevalence/incidence
Mason, 1993[171]	One-time ultrasound screening to men aged 70 years	Decision tree	5 cm	3.5–5 and >5 cm	100% sensitivity and specific	Prevalence rates: 5.3% (3.5–5 cm), 1.5% (>5 cm). Incidence not modelled
Sources			Based on aneurysm management study	Based on aneurysm management study	Not referenced	Not referenced
Frame et al., 1993[170]	One-time ultrasound or physical examination screening of men aged 60–80 years	20-year Markov model with annual cycles	4 cm	<4 and >4 cm	100% sensitivity and specific for ultrasound, 35% and 97% for physical examination, respectively	Prevalence rates: 3.1% (<4 cm), 2.3% (>4 cm). Incidence 0.1% per year
Sources			Based on available data, recognising that 5 cm is the common threshold	Based on available data describing rupture rates	Ultrasound sensitivity rates below 100% are referenced (with test results confirmed by surgery)	Prevalence based on screening studies of hypertensive patients and normal populations. Incidence rate is assumed
Law et al., 1994[181]	Non-specified frequency of ultrasound screening of men aged 60–80 years	Equation-based model[a]	6 cm	<3, –4, –5, –6 and >6 cm	86% sensitivity using a 6-cm threshold. The number of excess operations undertaken is estimated using the ratio of ruptured to unruptured aneurysms in 3 necropsy studies	Not relevant. Model uses AAA mortality rates directly
Sources			Based on sensitivity and specificity of different thresholds	Used solely to estimate sensitivity and specificity	Threshold detection rates based on smoothed (using cubic splines) data on distribution of ruptured aneurysm sizes in surgical and necropsy series	AAA-specific mortality rates obtained from Office of Population Censuses and Surveys
St Leger et al., 1996[174]	One-time ultrasound screening for AAA in men aged 65–74 years	5-year Markov model with annual cycles	6 cm	<3, –3.9, –4.9, –5.9 and >6 cm for prevalence	100% sensitivity and specific implied	Prevalence rates: 5.3% (3–3.9 cm), 1.4% (4–4.9 cm), 0.5% (5–5.9 cm), 0.6% (>6 cm)
Sources			Assumption	Implied to fit growth rates data	Not referenced	Based on various screening studies

TABLE 35 *Screening for abdominal aortic aneurysms modelling study characteristics (cont'd)*

Study	Comparators	Model type	Threshold for surgery	Aneurysm size	Test characteristics	Prevalence/incidence
Pentikainen et al., 2000[172]	One-time ultrasound screening for AAA for first-degree relatives, aged between 50 and 84 years	Lifetime (first-order?) simulation model	5 cm	2–2.5, –3, –5, >5 cm	100% sensitivity and specific implied	Prevalence rates: 13.4% (2–2.5 cm), 1.7% (2.5–3 cm), 1.7% (3–5 cm), 7.2% (>5 cm)
Sources					Based on two screening studies	Based on local data
Soisalon-Soininen et al., 2001[173]	One-time ultrasound screening for AAA for first-degree relatives, aged between 50–84 years	Decision model, same structure as Pentikainen et al.[172]	5 cm	2–2.5, –3, –5, >5 cm	100% sensitivity and specificity implied	Prevalence rates: 13.4% (2–2.5 cm), 1.7% (2.5–3 cm), 1.7% (3–5 cm), 7.2% (>5 cm)
Sources						Based on local data
Lee et al., 2002[182]	One-time ultrasound or 'quickscreen' screening for 70-year-old male patients with a high prevalence of AAA	Lifetime Markov model with annual cycles	5 cm	<3, –4, –5 and >5 cm	Both screening tests are 100% sensitivity and specificity	Prevalence rates: 3.01% (3–4 cm), 2.52% (4–5 cm), 1.47% (>5 cm)
Sources			"Available on request"	"Available on request"	Based on local study of 25 patients, in whom 7 AAAs were detected	"Available on request"
Mason, 1993[171]	20% of small aneurysms will not grow, 80% grow to >5 cm at 10% per year	Aneurysms >5 cm rupture at a rate of 5% per year	No interval cases or natural detection assumed	60% die before surgery, 50% die due to surgery. Total 80% mortality	25% contraindicated, 5% operative mortality	Survival curves for elective and emergency surgery survivors compared to normal all-cause mortality to estimate relative risk of death for the two groups. The relative risks are applied over a 10-year time horizon
Sources	Based on a study describing the natural history of AAAs	Based on a study describing prognosis of AAAs	Not referenced	Not referenced	Referenced	Based on prognostic factors study

continued

144

TABLE 35 *Screening for abdominal aortic aneurysms modelling study characteristics (cont'd)*

Study	Growth rates	Rupture rates	Interval and incidental presentations	Rupture mortality rates	Surgery contraindication and elective mortality	Post-surgery survival
Frame et al., 1993[170]	6% of aneurysms <4 cm become >4 cm per year	Aneurysms >4 cm rupture at a rate of 6% per year	No interval cases assumed. 33% incidental rate[b]	62% die before surgery, 50% die due to surgery. Total 81% mortality	30% contraindicated, 5% operative mortality	Assumed to be same as for normal population
Sources	Based on a study looking at surgical decisions following ultrasound	Reference missing	Not referenced	Based on single surgery impact study	Referenced	Not referenced
Law et al., 1994[181]	Not relevant. Model uses AAA mortality rates directly	Logistic regression estimated the risk of rupture by size, although not used in cost-effectiveness analysis	Not relevant. Screening effectiveness is based on aneurysm size at rupture, AAA mortality rates used directly	18% survive rupture	Not mentioned	Ratio of the ratio of circulatory to non-circulatory causes of death in post-AAA surgery survivors to that of the general population was applied to life expectancies for the normal population
Sources	Not relevant	13 studies describing ruptured aneurysms in patients in whom aneurysms were observed without intervention	Not relevant	Aggregation of 6 rupture outcome studies (5 UK studies)	–	Based on 15 US case series
St Leger et al., 1996[174]	Annual growth rates of 0.26, 0.46 and 0.66 cm for sizes 3–3.9, 4–4.9 and 5–5.9, respectively	Annual rupture rates of 0.6, 3.5 and 10% for sizes 4–4.9, 5–5.9 and >6 cm, respectively	Interval ruptures in aneurysms 4–4.9 and 5–5.9 cm modelled. No natural detection assumed	0% survive rupture	0% contraindicated, 5% operative mortality	Normal life expectancy estimates are applied to the model-estimated number of AAA-related deaths prevented
Sources	Based on various natural history studies	Based on various natural history studies	–	Based on same surgery impact study[c]	Operative mortality referenced	Based on single screening study

continued

TABLE 35 *Screening for abdominal aortic aneurysms modelling study characteristics (cont'd)*

Study	Growth rates	Rupture rates	Interval and incidental presentations	Rupture mortality rates	Surgery contraindication and elective mortality	Post-surgery survival
Pentikainen et al., 2000[172]	80% of small aneurysms will not grow. All other aneurysms grow at 5% per year	Quadratic increasing risk function: 0% risk at 4 cm, 5% annual risk at 5 cm and 100% risk at 10 cm	Interval ruptures in aneurysms 3–5 cm modelled. Natural detection is modelled but not described	60% die before surgery, 38% die due to surgery. Total 75.2% mortality	20% contraindicated, 4% operative mortality	Separate 15-year survival curves for elective and emergency patients
Sources	Two references "support" exponential growth, although limitations recognised	Distributional form is assumed, increasing rupture rates by size referenced	–	Implied based on Finnish data	Operative mortality based on Finnish data	Based on local data
Soisalon-Soininen et al., 2001[173]	80% of small aneurysms will not grow. All other aneurysms grow at 5% per year	Aneurysms >5 cm rupture at a rate of 5% per year	Interval ruptures in aneurysms 3–5 cm modelled. 63% natural detection for elective surgery	86% total mortality	20% contraindicated, 4% operative mortality	Separate 15-year survival curves for elective and emergency patients
Sources	Various studies referenced	Conservative assumption based on four studies	Incidental presentations based on local data	Based on local data	Operative mortality based on Finnish data	Based on local data
Lee et al., 2002[182]	Annual growth rates: 1 mm (3 cm), 2.6 mm (3–4 cm), 4 mm (4–5 cm)	Annual rupture rates: 0.1% (3 cm), 0.6% (3–4 cm), 2.3% (4–5 cm), 7.0% (>5 cm)	6.6% per year	30-day mortality 49%	0% contraindicated, 4% operative mortality	All patients undergoing surgery may develop the following complications: renal failure, stroke, MI or major amputation. An annual excess mortality risk is attributed to the first 3 complications
Sources	Described as weighted averages, 'available on request'	Described as weighted averages, 'available on request'	'Available on request'	Described as weighted averages, 'available on request'	Operative mortality described as weighted average, 'available on request'	Incidence of complications described as weighted averages, 'available on request'. Excess mortality risk referenced to three relevant studies

a Simple equation-based model, incorporating AAA deaths, threshold aneurysm size, sensitivity and specificity and ratio of ruptured to non-ruptured aneurysms.
b It is not clear how aneurysms are assumed to present; for example, do 33% of aneurysms reaching 4 cm present at that time?
c Same surgical intervention study as referenced by Frame and colleagues,[170] which shows overall mortality rate between 80 and 90%.

Health Technology Assessment reports published to date

Volume 1, 1997

No. 1
Home parenteral nutrition: a systematic review.
By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

No. 2
Diagnosis, management and screening of early localised prostate cancer.
A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

No. 3
The diagnosis, management, treatment and costs of prostate cancer in England and Wales.
A review by Chamberlain J, Melia J, Moss S, Brown J.

No. 4
Screening for fragile X syndrome.
A review by Murray J, Cuckle H, Taylor G, Hewison J.

No. 5
A review of near patient testing in primary care.
By Hobbs FDR, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al*.

No. 6
Systematic review of outpatient services for chronic pain control.
By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

No. 7
Neonatal screening for inborn errors of metabolism: cost, yield and outcome.
A review by Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al*.

No. 8
Preschool vision screening.
A review by Snowdon SK, Stewart-Brown SL.

No. 9
Implications of socio-cultural contexts for the ethics of clinical trials.
A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

No. 10
A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.
By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

No. 11
Newborn screening for inborn errors of metabolism: a systematic review.
By Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al*.

No. 12
Routine preoperative testing: a systematic review of the evidence.
By Munro J, Booth A, Nicholl J.

No. 13
Systematic review of the effectiveness of laxatives in the elderly.
By Petticrew M, Watt I, Sheldon T.

No. 14
When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies.
A review by Mowatt G, Bower DJ, Brebner JA, Cairns JA, Grant AM, McKee L.

Volume 2, 1998

No. 1
Antenatal screening for Down's syndrome.
A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

No. 2
Screening for ovarian cancer: a systematic review.
By Bell R, Petticrew M, Luengo S, Sheldon TA.

No. 3
Consensus development methods, and their use in clinical guideline development.
A review by Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al*.

No. 4
A cost–utility analysis of interferon beta for multiple sclerosis.
By Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D.

No. 5
Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.
By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al*.

No. 6
Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.
By Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G.

No. 7
Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.
By Song F, Glenny AM.

No. 8
Bone marrow and peripheral blood stem cell transplantation for malignancy.
A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

No. 9
Screening for speech and language delay: a systematic review of the literature.
By Law J, Boyle J, Harris F, Harkness A, Nye C.

No. 10
Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions.
By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

No. 11
Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.
By Ebrahim S.

No. 12
Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.
By McQuay HJ, Moore RA.

No. 13
Choosing between randomised and nonrandomised studies: a systematic review.
By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

No. 14
Evaluating patient-based outcome measures for use in clinical trials.
A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.

No. 15
Ethical issues in the design and conduct of randomised controlled trials.
A review by Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J.

No. 16
Qualitative research methods in health technology assessment: a review of the literature.
By Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P.

No. 17
The costs and benefits of paramedic skills in pre-hospital trauma care.
By Nicholl J, Hughes S, Dixon S, Turner J, Yates D.

No. 18
Systematic review of endoscopic ultrasound in gastro-oesophageal cancer.
By Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al*.

No. 19
Systematic reviews of trials and other studies.
By Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F.

No. 20
Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses.
A review by Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al*.

Volume 3, 1999

No. 1
Informed decision making: an annotated bibliography and systematic review.
By Bekker H, Thornton JG, Airey CM, Connelly JB, Hewison J, Robinson MB, *et al*.

No. 2
Handling uncertainty when performing economic evaluation of healthcare interventions.
A review by Briggs AH, Gray AM.

No. 3
The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.
By Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H.

No. 4
A randomised controlled trial of different approaches to universal antenatal HIV testing: uptake and acceptability. Annex: Antenatal HIV testing – assessment of a routine voluntary approach.
By Simpson WM, Johnstone FD, Boyd FM, Goldberg DJ, Hart GJ, Gormley SM, *et al*.

No. 5
Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.
By Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.

No. 6
Assessing the costs of healthcare technologies in clinical trials.
A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

No. 7
Cooperatives and their primary care emergency centres: organisation and impact.
By Hallam L, Henthorne K.

No. 8
Screening for cystic fibrosis.
A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

No. 9
A review of the use of health status measures in economic evaluation.
By Brazier J, Deverill M, Green C, Harper R, Booth A.

No. 10
Methods for the analysis of quality-of-life and survival data in health technology assessment.
A review by Billingham LJ, Abrams KR, Jones DR.

No. 11
Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.
By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

No. 12
Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses.
A review by Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al*.

No. 13
'Early warning systems' for identifying new healthcare technologies.
By Robert G, Stevens A, Gabbay J.

No. 14
A systematic review of the role of human papillomavirus testing within a cervical screening programme.
By Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, *et al*.

No. 15
Near patient testing in diabetes clinics: appraising the costs and outcomes.
By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

No. 16
Positron emission tomography: establishing priorities for health technology assessment.
A review by Robert G, Milne R.

No. 17 (Pt 1)
The debridement of chronic wounds: a systematic review.
By Bradley M, Cullum N, Sheldon T.

No. 17 (Pt 2)
Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.
By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

No. 18
A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.
By Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al*.

No. 19
What role for statins? A review and economic model.
By Ebrahim S, Davey Smith G, McCabe C, Payne N, Pickin M, Sheldon TA, *et al*.

No. 20
Factors that limit the quality, number and progress of randomised controlled trials.
A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, *et al*.

No. 21
Antimicrobial prophylaxis in total hip replacement: a systematic review.
By Glenny AM, Song F.

No. 22
Health promoting schools and health promotion in schools: two systematic reviews.
By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

No. 23
Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.
A review by Lord J, Victor C, Littlejohns P, Ross FM, Axford JS.

Volume 4, 2000

No. 1
The estimation of marginal time preference in a UK-wide sample (TEMPUS) project.
A review by Cairns JA, van der Pol MM.

No. 2
Geriatric rehabilitation following fractures in older people: a systematic review.
By Cameron I, Crotty M, Currie C, Finnegan T, Gillespie L, Gillespie W, *et al*.

No. 3
Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.
By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

No. 4
Community provision of hearing aids and related audiology services.
A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

No. 5
False-negative results in screening programmes: systematic review of impact and implications.
By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

No. 6
Costs and benefits of community postnatal support workers: a randomised controlled trial.
By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

No. 7
Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.
By French RS, Cowan FM, Mansour DJA, Morris S, Procter T, Hughes D, *et al*.

No. 8
An introduction to statistical methods for health technology assessment.
A review by White SJ, Ashby D, Brown PJ.

No. 9
Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.
By Clegg A, Bryant J, Milne R.

No. 10
Publication and related biases.
A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

No. 11
Cost and outcome implications of the organisation of vascular services.
By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

No. 12
Monitoring blood glucose control in diabetes mellitus: a systematic review.
By Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R.

No. 13
The effectiveness of domiciliary health visiting: a systematic review of international studies
and a selective review of the British literature.
By Elkan R, Kendrick D, Hewitt M, Robinson JJA, Tolley K, Blair M, *et al*.

No. 14
The determinants of screening uptake and interventions for increasing uptake: a systematic review.
By Jepson R, Clegg A, Forbes C, Lewis R, Sowden A, Kleijnen J.

No. 15
The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth.
A rapid review by Song F, O'Meara S, Wilson P, Golder S, Kleijnen J.

No. 16
Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views.
By Bricker L, Garcia J, Henderson J, Mugford M, Neilson J, Roberts T, *et al*.

No. 17
A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.
By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

No. 18
Liquid-based cytology in cervical screening: a rapid and systematic review.
By Payne N, Chilcott J, McGoogan E.

No. 19
Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.
By King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, *et al*.

No. 20
Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography?
By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

No. 21
Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.
By O'Meara S, Cullum N, Majid M, Sheldon T.

No. 22
Using routine data to complement and enhance the results of randomised controlled trials.
By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

No. 23
Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.
By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

No. 24
Outcome measures for adult critical care: a systematic review.
By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al*.

No. 25
A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.
By Fairbank L, O'Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

No. 26
Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.
By Parkes J, Bryant J, Milne R.

No. 27
Treatments for fatigue in multiple sclerosis: a rapid and systematic review.
By Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

No. 28
Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.
By Baxter-Jones ADG, Helms PJ, Russell G, Grant A, Ross S, Cairns JA, *et al*.

No. 29
Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.
By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

No. 30
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.
By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

No. 31
A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.
By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

No. 32
Intrathecal pumps for giving opioids in chronic pain: a systematic review.
By Williams JE, Louw G, Towlerton G.

No. 33
Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.
By Shepherd J, Waugh N, Hewitson P.

149

No. 34
A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.
By MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

No. 35
Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.
By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, *et al*.

No. 36
A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.
By Simpson S, Corney R, Fitzgerald P, Beecham J.

No. 37
Systematic review of treatments for atopic eczema.
By Hoare C, Li Wan Po A, Williams H.

No. 38
Bayesian methods in health technology assessment: a review.
By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

No. 39
The management of dyspepsia: a systematic review.
By Delaney B, Moayyedi P, Deeks J, Innes M, Soo S, Barton P, *et al*.

No. 40
A systematic review of treatments for severe psoriasis.
By Griffiths CEM, Clark CM, Chalmers RJG, Li Wan Po A, Williams HC.

Volume 5, 2001

No. 1
Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer's disease: a rapid and systematic review.
By Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K, *et al*.

No. 2
The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.
By Stewart A, Sandercock J, Bryan S, Hyde C, Barton PM, Fry-Smith A, *et al*.

No. 3
Equity and the economic evaluation of healthcare.
By Sassi F, Archard L, Le Grand J.

No. 4
Quality-of-life measures in chronic diseases of childhood.
By Eiser C, Morse R.

No. 5
Eliciting public preferences for healthcare: a systematic review of techniques.
By Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al*.

No. 6
General health status measures for people with cognitive impairment: learning disability and acquired brain injury.
By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

No. 7
An assessment of screening strategies for fragile X syndrome in the UK.
By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

No. 8
Issues in methodological research: perspectives from researchers and commissioners.
By Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, *et al*.

No. 9
Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.
By Cullum N, Nelson EA, Flemming K, Sheldon T.

No. 10
Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.
By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, *et al*.

No. 11
Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.
By Jobanputra P, Parry D, Fry-Smith A, Burls A.

No. 12
Statistical assessment of the learning curves of health technologies.
By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

No. 13
The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.
By Dinnes J, Cave C, Huang S, Major K, Milne R.

No. 14
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.
By Lewis R, Whiting P, ter Riet G, O'Meara S, Glanville J.

No. 15
Home treatment for mental health problems: a systematic review.
By Burns T, Knapp M, Catty J, Healey A, Henderson J, Watt H, *et al*.

No. 16
How to develop cost-conscious guidelines.
By Eccles M, Mason J.

No. 17
The role of specialist nurses in multiple sclerosis: a rapid and systematic review.
By De Broe S, Christopher F, Waugh N.

No. 18
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.
By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 19
The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.
By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

No. 20
Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in pre-operative assessment in elective general surgery.
By Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, *et al*.

No. 21
Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.
By Marshall M, Crowther R, Almaraz-Serrano A, Creed F, Sledge W, Kluiter H, *et al*.

No. 22
The measurement and monitoring of surgical adverse events.
By Bruce J, Russell EM, Mollison J, Krukowski ZH.

No. 23
Action research: a systematic review and guidance for assessment.
By Waterman H, Tillen D, Dickson R, de Koning K.

No. 24
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.
By Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al*.

No. 25
A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.
By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

No. 26
Comparison of the effectiveness of inhaler devices in asthma and chronic obstructive airways disease: a systematic review of the literature.
By Brocklebank D, Ram F, Wright J, Barry P, Cates C, Davies L, *et al.*

No. 27
The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.
By Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al.*

No. 28
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.
By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

No. 29
Superseded by a report published in a later volume.

No. 30
The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.
By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

No. 31
Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.
By McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al.*

No. 32
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.
By Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N.

No. 33
Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.
By Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

No. 34
Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.
By David AS, Adams C.

No. 35
A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.
By Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al.*

No. 36
Cost analysis of child health surveillance.
By Sanderson D, Wright D, Acton C, Duree D.

Volume 6, 2002

No. 1
A study of the methods used to select review criteria for clinical audit.
By Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

No. 2
Fludarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.
By Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al.*

No. 3
Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation.
By Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al.*

No. 4
A systematic review of discharge arrangements for older people.
By Parker SG, Peet SM, McPherson A, Cannaby AM, Baker R, Wilson A, *et al.*

No. 5
The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.
By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

No. 6
The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.
By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 7
The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.
By Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, *et al.*

No. 8
Promoting physical activity in South Asian Muslim women through 'exercise on prescription'.
By Carroll B, Ali N, Azam N.

No. 9
Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.
By Burls A, Clark W, Stewart T, Preston C, Bryan S, Jefferson T, *et al.*

No. 10
A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.
By Richards RG, Sampson FC, Beard SM, Tappenden P.

No. 11
Screening for gestational diabetes: a systematic review and economic evaluation.
By Scott DA, Loveman E, McIntyre L, Waugh N.

No. 12
The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.
By Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A.

No. 13
The clinical effectiveness of trastuzumab for breast cancer: a systematic review.
By Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al.*

No. 14
The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.
By Lewis R, Bagnall A-M, King S, Woolacott N, Forbes C, Shirran L, *et al.*

No. 15
A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease.
By Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC.

No. 16
The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation.
By Woolacott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, *et al.*

No. 17
A systematic review of effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept.
By Cummins C, Connock M, Fry-Smith A, Burls A.

No. 18
Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation.
By Bryant J, Cave C, Mihaylova B, Chase D, McIntyre L, Gerard K, *et al.*

151

No. 19
Clinical effectiveness and cost-effectiveness of growth hormone in adults in relation to impact on quality of life: a systematic review and economic evaluation.
By Bryant J, Loveman E, Chase D, Mihaylova B, Cave C, Gerard K, *et al*.

No. 20
Clinical medication review by a pharmacist of patients on repeat prescriptions in general practice: a randomised controlled trial.
By Zermansky AG, Petty DR, Raynor DK, Lowe CJ, Freementle N, Vail A.

No. 21
The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.
By Jobanputra P, Barton P, Bryan S, Burls A.

No. 22
A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.
By Kaltenthaler E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J.

No. 23
A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer.
By Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Reimsma R.

No. 24
A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change.
By Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, *et al*.

No. 25
A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists.
By Robinson M, Ginnelly L, Sculpher M, Jones L, Riemsma R, Palmer S, *et al*.

No. 26
A systematic review of the effectiveness, cost-effectiveness and barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS.
By Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, *et al*.

No. 27
A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic.
By Caine N, Sharples LD, Hollingworth W, French J, Keogan M, Exley A, *et al*.

No. 28
Clinical effectiveness and cost – consequences of selective serotonin reuptake inhibitors in the treatment of sex offenders.
By Adi Y, Ashcroft D, Browne K, Beech A, Fry-Smith A, Hyde C.

No. 29
Treatment of established osteoporosis: a systematic review and cost–utility analysis.
By Kanis JA, Brazier JE, Stevenson M, Calvert NW, Lloyd Jones M.

No. 30
Which anaesthetic agents are cost-effective in day surgery? Literature review, national survey of practice and randomised controlled trial.
By Elliott RA Payne K, Moore JK, Davies LM, Harper NJN, St Leger AS, *et al*.

No. 31
Screening for hepatitis C among injecting drug users and in genitourinary medicine clinics: systematic reviews of effectiveness, modelling study and national survey of current practice.
By Stein K, Dalziel K, Walker A, McIntyre L, Jenkins B, Horne J, *et al*.

No. 32
The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.
By Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, *et al*.

No. 33
The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review.
By Garside R, Round A, Dalziel K, Stein K, Royle R.

No. 34
A comparative study of hypertonic saline, daily and alternate-day rhDNase in children with cystic fibrosis.
By Suri R, Wallis C, Bush A, Thompson S, Normand C, Flather M, *et al*.

No. 35
A systematic review of the costs and effectiveness of different models of paediatric home care.
By Parker G, Bhakta P, Lovett CA, Paisley S, Olsen R, Turner D, *et al*.

Volume 7, 2003

No. 1
How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.
By Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J.

No. 2
Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure.
By Mowatt G, Vale L, Perez J, Wyness L, Fraser C, MacLeod A, *et al*.

No. 3
Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease.
By Clark W, Raftery J, Barton P, Song F, Fry-Smith A, Burls A.

No. 4
A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus negative.
By Chilcott J, Lloyd Jones M, Wight J, Forman K, Wray J, Beverley C, *et al*.

No. 5
Systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma.
By Riley RD, Burchill SA, Abrams KR, Heney D, Lambert PC, Jones DR, *et al*.

No. 6
The cost-effectiveness of screening for *Helicobacter pylori* to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete-event simulation model.
By Roderick P, Davies R, Raftery J, Crabbe D, Pearce R, Bhandari P, *et al*.

No. 7
The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation.
By Davenport C, Elley K, Salas C, Taylor-Weetman CL, Fry-Smith A, Bryan S, *et al*.

No. 8
A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia.
By Kennedy ADM, Sculpher MJ, Coulter A, Dwyer N, Rees M, Horsley S, *et al*.

No. 9
Clinical effectiveness and cost–utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation.
By Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde C.

No. 10
Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.
By Grimshaw GM, Szczepura A, Hultén M, MacDonald F, Nevin NC, Sutton F, *et al*.

No. 11
First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS).
By Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM.

No. 12
The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation.
By Calvert N, Hind D, McWilliams RG, Thomas SM, Beverley C, Davidson A.

No. 13
A systematic review of atypical antipsychotics in schizophrenia.
By Bagnall A-M, Jones L, Lewis R, Ginnelly L, Glanville J, Torgerson D, *et al*.

No. 14
Prostate Testing for Cancer and Treatment (ProtecT) feasibility study.
By Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, *et al*.

No. 15
Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation.
By Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica Mota R, *et al*.

No. 16
Screening for fragile X syndrome: a literature review and modelling.
By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

No. 17
Systematic review of endoscopic sinus surgery for nasal polyps.
By Dalziel K, Stein K, Round A, Garside R, Royle P.

No. 18
Towards efficient guidelines: how to monitor guideline use in primary care.
By Hutchinson A, McIntosh A, Cox S, Gilbert C.

No. 19
Effectiveness and cost-effectiveness of acute hospital-based spinal cord injuries services: systematic review.
By Bagnall A-M, Jones L, Richardson G, Duffy S, Riemsma R.

No. 20
Prioritisation of health technology assessment. The PATHS model: methods and case studies.
By Townsend J, Buxton M, Harper G.

No. 21
Systematic review of the clinical effectiveness and cost-effectiveness of tension-free vaginal tape for treatment of urinary stress incontinence.
By Cody J, Wyness L, Wallace S, Glazener C, Kilonzo M, Stearns S, *et al*.

No. 22
The clinical and cost-effectiveness of patient education models for diabetes: a systematic review and economic evaluation.
By Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N.

No. 23
The role of modelling in prioritising and planning clinical trials.
By Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P.

No. 24
Cost–benefit evaluation of routine influenza immunisation in people 65–74 years of age.
By Allsup S, Gosney M, Haycox A, Regan M.

No. 25
The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors.
By Wight J, Chilcott J, Holmes M, Brewer N.

No. 26
Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment.
By Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT.

No. 27
Evaluating non-randomised intervention studies.
By Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*.

No. 28
A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease.
By Kennedy A, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, *et al*.

No. 29
The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review.
By Dinnes J, Loveman E, McIntyre L, Waugh N.

No. 30
The value of digital imaging in diabetic retinopathy.
By Sharp PF, Olson J, Strachan F, Hipwell J, Ludbrook A, O'Donnell M, *et al*.

No. 31
Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy.
By Law M, Wald N, Morris J.

No. 32
Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: systematic review and economic evaluation.
By Ward S, Kaltenthaler E, Cowan J, Brewer N.

No. 33
Clinical and cost-effectiveness of new and emerging technologies for early localised prostate cancer: a systematic review.
By Hummel S, Paisley S, Morgan A, Currie E, Brewer N.

No. 34
Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.
By Royle P, Waugh N.

No. 35
Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.
By Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K.

No. 36
A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses.
By Boland A, Haycox A, Bagust A, Fitzsimmons L.

No. 37
Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women's physical and psychological health needs.
By MacArthur C, Winter HR, Bick DE, Lilford RJ, Lancashire RJ, Knowles H, *et al*.

No. 38
Estimating implied rates of discount in healthcare decision-making.
By West RR, McNabb R, Thompson AGH, Sheldon TA, Grimley Evans J.

153

No. 39
Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling.
By Cooper BS, Stone SP, Kibbler CC, Cookson BD, Roberts JA, Medley GF, *et al*.

No. 40
Treatments for spasticity and pain in multiple sclerosis: a systematic review.
By Beard S, Hunn A, Wight J.

No. 41
The inclusion of reports of randomised trials published in languages other than English in systematic reviews.
By Moher D, Pham B, Lawson ML, Klassen TP.

No. 42
The impact of screening on future health-promoting behaviours and health beliefs: a systematic review.
By Bankhead CR, Brett J, Bukach C, Webster P, Stewart-Brown S, Munafo M, *et al*.

Volume 8, 2004

No. 1
What is the best imaging strategy for acute stroke?
By Wardlaw JM, Keir SL, Seymour J, Lewis S, Sandercock PAG, Dennis MS, *et al*.

No. 2
Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care.
By Mant J, McManus RJ, Oakes RAL, Delaney BC, Barton PM, Deeks JJ, *et al*.

No. 3
The effectiveness and cost-effectiveness of microwave and thermal balloon endometrial ablation for heavy menstrual bleeding: a systematic review and economic modelling.
By Garside R, Stein K, Wyatt K, Round A, Price A.

No. 4
A systematic review of the role of bisphosphonates in metastatic disease.
By Ross JR, Saunders Y, Edmonds PM, Patel S, Wonderling D, Normand C, *et al*.

No. 5
Systematic review of the clinical effectiveness and cost-effectiveness of capecitabine (Xeloda®) for locally advanced and/or metastatic breast cancer.
By Jones L, Hawkins N, Westwood M, Wright K, Richardson G, Riemsma R.

No. 6
Effectiveness and efficiency of guideline dissemination and implementation strategies.
By Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al*.

No. 7
Clinical effectiveness and costs of the Sugarbaker procedure for the treatment of pseudomyxoma peritonei.
By Bryant J, Clegg AJ, Sidhu MK, Brodin H, Royle P, Davidson P.

No. 8
Psychological treatment for insomnia in the regulation of long-term hypnotic drug use.
By Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M.

No. 9
Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome.
By Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ.

No. 10
A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography.
By Kaltenthaler E, Bravo Vergel Y, Chilcott J, Thomas S, Blakeborough T, Walters SJ, *et al*.

No. 11
The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis.
By Barton P, Jobanputra P, Wilson J, Bryan S, Burls A.

No. 12
Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review.
By Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S.

No. 13
Clinical effectiveness and cost-effectiveness of pioglitazone and rosiglitazone in the treatment of type 2 diabetes: a systematic review and economic evaluation.
By Czoski-Murray C, Warren E, Chilcott J, Beverley C, Psyllaki MA, Cowan J.

No. 14
Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers.
By Townsend J, Wolke D, Hayes J, Davé S, Rogers C, Bloomfield L, *et al*.

No. 15
Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.
By Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, *et al*.

No. 16
A multi-centre randomised controlled trial of minimally invasive direct coronary bypass grafting versus percutaneous transluminal coronary angioplasty with stenting for proximal stenosis of the left anterior descending coronary artery.
By Reeves BC, Angelini GD, Bryan AJ, Taylor FC, Cripps T, Spyt TJ, *et al*.

No. 17
Does early magnetic resonance imaging influence management or improve outcome in patients referred to secondary care with low back pain? A pragmatic randomised controlled trial.
By Gilbert FJ, Grant AM, Gillan MGC, Vale L, Scott NW, Campbell MK, *et al*.

No. 18
The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis.
By Clark W, Jobanputra P, Barton P, Burls A.

No. 19
A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.
By Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, *et al*.

No. 20
Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis.
By Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N.

No. 21
Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.
By Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, *et al*.

No. 22
Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus.
By Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A.

No. 23
Clinical effectiveness and cost-effectiveness of prehospital intravenous fluids in trauma patients.
By Dretzke J, Sandercock J, Bayliss S, Burls A.

No. 24
Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation.
By Dündar Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust A, *et al*.

No. 25
Development and validation of methods for assessing the quality of diagnostic accuracy studies.
By Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J.

No. 26
EVALUATE hysterectomy trial: a multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy.
By Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al*.

No. 27
Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis.
By Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C.

No. 28
Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis.
By Dalziel K, Round A, Stein K, Garside R, Price A.

No. 29
VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.
By Iglesias C, Nelson EA, Cullum NA, Torgerson DJ on behalf of the VenUS Team.

No. 30
Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction.
By Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, *et al*.

No. 31
A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme.
By Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S.

No. 32
The Social Support and Family Health Study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner-city areas.
By Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, *et al*.

No. 33
Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review.
By Green JM, Hewison J, Bekker HL, Bryant, Cuckle HS.

No. 34
Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status.
By Critchley HOD, Warner P, Lee AJ, Brechin S, Guise J, Graham B.

No. 35
Coronary artery stents: a rapid systematic review and economic evaluation.
By Hill R, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, *et al*.

No. 36
Review of guidelines for good practice in decision-analytic modelling in health technology assessment.
By Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*.

No. 37
Rituximab (MabThera®) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation.
By Knight C, Hind D, Brewer N, Abbott V.

No. 38
Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation.
By Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al*.

No. 39
Pegylated interferon α-2a and -2b in combination with ribavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation.
By Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J.

No. 40
Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation.
By Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al*.

No. 41
Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups.
By Beswick AD, Rees K, Griebsch I, Taylor FC, Burke M, West RR, *et al*.

No. 42
Involving South Asian patients in clinical trials.
By Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P.

No. 43
Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes.
By Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N.

No. 44
Identification and assessment of ongoing trials in health technology assessment reviews.
By Song FJ, Fry-Smith A, Davenport C, Bayliss S, Adi Y, Wilson JS, *et al*.

No. 45
Systematic review and economic evaluation of a long-acting insulin analogue, insulin glargine
By Warren E, Weatherley-Jones E, Chilcott J, Beverley C.

No. 46
Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis.
By McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al*.

No. 47
Clinical and cost-effectiveness of once-daily versus more frequent use of same potency topical corticosteroids for atopic eczema: a systematic review and economic evaluation.
By Green C, Colquitt JL, Kirby J, Davidson P, Payne E.

No. 48
Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis.
By Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al*.

No. 49
Generalisability in economic evaluation studies in healthcare: a review and case studies.
By Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al*.

No. 50
Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations.
By Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, *et al*.

155

No. 28
Outcomes of electrically stimulated gracilis neosphincter surgery.
By Tillin T, Chambers M, Feldman R.

No. 29
The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation.
By Garside R, Stein K, Castelnuovo E, Pitt M, Ashcroft D, Dimmock P, *et al.*

No. 30
Systematic review on urine albumin testing for early detection of diabetic complications.
By Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, *et al.*

No. 31
Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis.
By Cochrane T, Davey RC, Matthes Edwards SM.

No. 32
Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain.
By Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, *et al.*

No. 33
Cost-effectiveness and safety of epidural steroids in the management of sciatica.
By Price C, Arden N, Coglan L, Rogers P.

No. 34
The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis.
By Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL.

No. 35
Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials.
By King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al.*

No. 36
The clinical and cost-effectiveness of implantable cardioverter defibrillators: a systematic review.
By Bryant J, Brodin H, Loveman E, Payne E, Clegg A.

No. 37
A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study.
By Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, *et al.*

No. 38
The causes and effects of socio-demographic exclusions from clinical trials.
By Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, *et al.*

No. 39
Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis.
By Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, *et al.*

No. 40
A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study.
By Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al.*

No. 41
Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty.
By Keating JF, Grant A, Masson M, Scott NW, Forbes JF.

No. 42
Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland.
By Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, *et al.*

No. 43
The effectiveness and cost-effectiveness of dual-chamber pacemakers compared with single-chamber pacemakers for bradycardia due to atrioventricular block or sick sinus syndrome: systematic review and economic evaluation.
By Castelnuovo E, Stein K, Pitt M, Garside R, Payne E.

No. 44
Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis.
By Knowles R, Griebsch I, Dezateux C, Brown J, Bull C, Wren C.

No. 45
The clinical and cost-effectiveness of left ventricular assist devices for end-stage heart failure: a systematic review and economic evaluation.
By Clegg AJ, Scott DA, Loveman E, Colquitt J, Hutchinson J, Royle P, *et al.*

No. 46
The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma.
By Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D.

No. 47
Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation.
By Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al.*

No. 48
Systematic review of effectiveness of different treatments for childhood retinoblastoma.
By McDaid C, Hartley S, Bagnall A-M, Ritchie G, Light K, Riemsma R.

No. 49
Towards evidence-based guidelines for the prevention of venous thromboembolism: systematic reviews of mechanical methods, oral anticoagulation, dextran and regional anaesthesia as thromboprophylaxis.
By Roderick P, Ferris G, Wilson K, Halls H, Jackson D, Collins R, *et al.*

No. 50
The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children.
By Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, *et al.*

Volume 10, 2006

No. 1
The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease.
By Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, *et al.*

No. 2
FOOD: a multicentre randomised trial evaluating feeding policies in patients admitted to hospital with a recent stroke.
By Dennis M, Lewis S, Cranswick G, Forbes J.

No. 3
The clinical effectiveness and cost-effectiveness of computed tomography screening for lung cancer: systematic reviews.
By Black C, Bagust A, Boland A, Walker S, McLeod C, De Verteuil R, *et al.*

157

No. 4
A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery.
By Whiting P, Gupta R, Burch J, Mujica Mota RE, Wright K, Marson A, *et al*.

No. 5
Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies.
By Dundar Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR.

No. 6
Systematic review and evaluation of methods of assessing urinary incontinence.
By Martin JL, Williams KS, Abrams KR, Turner DA, Sutton AJ, Chapple C, *et al*.

No. 7
The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review.
By Connock M, Frew E, Evans B-W, Bryan S, Cummins C, Fry-Smith A, *et al*.

No. 8
Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling.
By Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N.

No. 9
Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation.
By Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, *et al*.

No. 10
Evaluation of molecular techniques in prediction and diagnosis of cytomegalovirus disease in immunocompromised patients.
By Szczepura A, Westmoreland D, Vinogradova Y, Fox J, Clark M.

No. 11
Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study.
By Wu O, Robertson L, Twaddle S, Lowe GDO, Clark P, Greaves M, *et al*.

No. 12
A series of systematic reviews to inform a decision analysis for sampling and treating infected diabetic foot ulcers.
By Nelson EA, O'Meara S, Craig D, Iglesias C, Golder S, Dalton J, *et al*.

No. 13
Randomised clinical trial, observational study and assessment of cost-effectiveness of the treatment of varicose veins (REACTIV trial).
By Michaels JA, Campbell WB, Brazier JE, MacIntyre JB, Palfreyman SJ, Ratcliffe J, *et al*.

No. 14
The cost-effectiveness of screening for oral cancer in primary care.
By Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M *et al*.

No. 15
Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis.
By Goodacre S, Sampson F, Stevenson M, Wailoo A, Sutton A, Thomas S, *et al*.

No. 16
Systematic review of the effectiveness and cost-effectiveness of HealOzone® for the treatment of occlusal pit/fissure caries and root caries.
By Brazzelli M, McKenzie L, Fielding S, Fraser C, Clarkson J, Kilonzo M, *et al*.

No. 17
Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment.
By Lewis SW, Davies L, Jones PB, Barnes TRE, Murray RM, Kerwin R, *et al*.

No. 18
Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation.
By Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, *et al*.

No. 19
Cognitive behavioural therapy in addition to antispasmodic therapy for irritable bowel syndrome in primary care: randomised controlled trial.
By Kennedy TM, Chalder T, McCrone P, Darnley S, Knapp M, Jones RH, *et al*.

No. 20
A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type 1.
By Connock M, Juarez-Garcia A, Frew E, Mans A, Dretzke J, Fry-Smith A, *et al*.

No. 21
Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation.
By Wright M, Grieve R, Roberts J, Main J, Thomas HC on behalf of the UK Mild Hepatitis C Trial Investigators.

No. 22
Pressure relieving support surfaces: a randomised evaluation.
By Nixon J, Nelson EA, Cranny G, Iglesias CP, Hawkins K, Cullum NA, *et al*.

No. 23
A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents.
By King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al*.

No. 24
The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review.
By Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, *et al*.

No. 25
Effectiveness and cost-effectiveness of salicylic acid and cryotherapy for cutaneous warts. An economic decision model.
By Thomas KS, Keogh-Brown MR, Chalmers JR, Fordham RJ, Holland RC, Armstrong SJ, *et al*.

No. 26
A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use.
By Robinson L, Hutchings D, Corner L, Beyer F, Dickinson H, Vanoli A, *et al*.

No. 27
A review of the evidence on the effects and costs of implantable cardioverter defibrillator therapy in different patient groups, and modelling of cost-effectiveness and cost–utility for these groups in a UK context.
By Buxton M, Caine N, Chase D, Connelly D, Grace A, Jackson C, *et al*.

No. 28
Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation.
By Shepherd J, Jones J, Takeda A, Davidson P, Price A.

No. 29
An evaluation of the clinical and cost-effectiveness of pulmonary artery catheters in patient management in intensive care: a systematic review and a randomised controlled trial.
By Harvey S, Stevens K, Harrison D, Young D, Brampton W, McCabe C, *et al*.

No. 30
Accurate, practical and cost-effective assessment of carotid stenosis in the UK.
By Wardlaw JM, Chappell FM, Stevenson M, De Nigris E, Thomas S, Gillard J, *et al*.

No. 31
Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.
By Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al*.

No. 32
The cost-effectiveness of testing for hepatitis C in former injecting drug users.
By Castelnuovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al*.

No. 33
Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.
By Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, *et al*.

No. 34
Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy.
By Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al*.

No. 35
Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation.
By Brazier J, Tumur I, Holmes M, Ferriter M, Parry G, Dent-Brown K, *et al*.

No. 36
Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model.
By Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al*.

No. 37
Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme.
By O'Dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A.

No. 38
A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling.
By Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, *et al*.

No. 39
The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review.
By Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G.

No. 40
What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? A Multi-Institution Nurse Endoscopy Trial (MINuET).
By Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al*.

No. 41
The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation.
By Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P.

No. 42
A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.
By Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al*.

No. 43
Telemedicine in dermatology: a randomised controlled trial.
By Bowns IR, Collins K, Walters SJ, McDonagh AJG.

No. 44
Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: a systematic review and economic model.
By Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C.

No. 45
Clinical effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer: systematic reviews and economic evaluation.
By Murray A, Lourenco T, de Verteuil R, Hernandez R, Fraser C, McKinley A, *et al*.

No. 46
Etanercept and efalizumab for the treatment of psoriasis: a systematic review.
By Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, *et al*.

No. 47
Systematic reviews of clinical decision tools for acute abdominal pain.
By Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al*.

No. 48
Evaluation of the ventricular assist device programme in the UK.
By Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, *et al*.

No. 49
A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children.
By Yao G, Albon E, Adi Y, Milford D, Bayliss S, Ready A, *et al*.

No. 50
Amniocentesis results: investigation of anxiety. The ARIA trial.
By Hewison J, Nixon J, Fountain J, Cocks K, Jones C, Mason G, *et al*.

Volume 11, 2007

No. 1
Pemetrexed disodium for the treatment of malignant pleural mesothelioma: a systematic review and economic evaluation.
By Dundar Y, Bagust A, Dickson R, Dodd S, Green J, Haycox A, *et al*.

No. 2
A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer.
By Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K, *et al*.

No. 3
A systematic review of rapid diagnostic tests for the detection of tuberculosis infection.
By Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, Waugh N, *et al*.

No. 4
The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women.
By Stevenson M, Davis S, Lloyd-Jones M, Beverley C.

159

No. 5
A systematic review of quantitative and qualitative research on the role and effectiveness of written information available to patients about individual medicines.

By Raynor DK, Blenkinsopp A, Knapp P, Grime J, Nicolson DJ, Pollock K, *et al.*

No. 6
Oral naltrexone as a treatment for relapse prevention in formerly opioid-dependent drug users: a systematic review and economic evaluation.

By Adi Y, Juarez-Garcia A, Wang D, Jowett S, Frew E, Day E, *et al.*

No. 7
Glucocorticoid-induced osteoporosis: a systematic review and cost–utility analysis.

By Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M.

No. 8
Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection.

By Low N, McCarthy A, Macleod J, Salisbury C, Campbell R, Roberts TE, *et al.*

No. 9
Methadone and buprenorphine for the management of opioid dependence: a systematic review and economic evaluation.

By Connock M, Juarez-Garcia A, Jowett S, Frew E, Liu Z, Taylor RJ, *et al.*

No. 10
Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only.

By Isaacs AJ, Critchley JA, See Tai S, Buckingham K, Westley D, Harridge SDR, *et al.*

No. 11
Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: a systematic review and economic evaluation.

By Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N.

No. 12
Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer.

By Tappenden P, Jones R, Paisley S, Carroll C.

No. 13
A systematic review and economic evaluation of epoetin alfa, epoetin beta and darbepoetin alfa in anaemia associated with cancer, especially that attributable to cancer treatment.

By Wilson J, Yao GL, Raftery J, Bohlius J, Brunskill S, Sandercock J, *et al.*

No. 14
A systematic review and economic evaluation of statins for the prevention of coronary events.

By Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al.*

No. 15
A systematic review of the effectiveness and cost-effectiveness of different models of community-based respite care for frail older people and their carers.

By Mason A, Weatherly H, Spilsbury K, Arksey H, Golder S, Adamson J, *et al.*

No. 16
Additional therapy for young children with spastic cerebral palsy: a randomised controlled trial.

By Weindling AM, Cunningham CC, Glenn SM, Edwards RT, Reeves DJ.

No. 17
Screening for type 2 diabetes: literature review and economic modelling.

By Waugh N, Scotland G, McNamee P, Gillett M, Brennan A, Goyder E, *et al.*

No. 18
The effectiveness and cost-effectiveness of cinacalcet for secondary hyperparathyroidism in end-stage renal disease patients on dialysis: a systematic review and economic evaluation.

By Garside R, Pitt M, Anderson R, Mealing S, Roome C, Snaith A, *et al.*

No. 19
The clinical effectiveness and cost-effectiveness of gemcitabine for metastatic breast cancer: a systematic review and economic evaluation.

By Takeda AL, Jones J, Loveman E, Tan SC, Clegg AJ.

No. 20
A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease.

By Collins R, Cranny G, Burch J, Aguiar-Ibáñez R, Craig D, Wright K, *et al.*

No. 21
The clinical effectiveness and cost-effectiveness of treatments for children with idiopathic steroid-resistant nephrotic syndrome: a systematic review.

By Colquitt JL, Kirby J, Green C, Cooper K, Trompeter RS.

No. 22
A systematic review of the routine monitoring of growth in children of primary school age to identify growth-related conditions.

By Fayter D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, *et al.*

No. 23
Systematic review of the effectiveness of preventing and treating *Staphylococcus aureus* carriage in reducing peritoneal catheter-related infections.

By McCormack K, Rabindranath K, Kilonzo M, Vale L, Fraser C, McIntyre L, *et al.*

No. 24
The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis.

By McLoughlin DM, Mogg A, Eranti S, Pluck G, Purvis R, Edwards D, *et al.*

No. 25
A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment.

By Boyle J, McCartney E, Forbes J, O'Hare A.

No. 26
Hormonal therapies for early breast cancer: systematic review and economic evaluation.

By Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L.

No. 27
Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review.

By Bryant J, Picot J, Levitt G, Sullivan I, Baxter L, Clegg A.

No. 28
Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation.

By McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, *et al.*

No. 29
Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses.
By Colbourn T, Asseburg C, Bojke L, Philips Z, Claxton K, Ades AE, *et al*.

No. 30
Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review.
By Garrison KR, Donell S, Ryder J, Shemilt I, Mugford M, Harvey I, *et al*.

No. 31
A randomised controlled trial of postoperative radiotherapy following breast-conserving surgery in a minimum-risk older population. The PRIME trial.
By Prescott RJ, Kunkler IH, Williams LJ, King CC, Jack W, van der Pol M, *et al*.

No. 32
Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen.
By Bamford J, Fortnum H, Bristow K, Smith J, Vamvakas G, Davies L, *et al*.

No. 33
The clinical effectiveness and cost-effectiveness of inhaled insulin in diabetes mellitus: a systematic review and economic evaluation.
By Black C, Cummins E, Royle P, Philip S, Waugh N.

No. 34
Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis.
By Thompson Coon J, Rogers G, Hewson P, Wright D, Anderson R, Cramp M, *et al*.

No. 35
The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Home-based compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence.
By Jolly K, Taylor R, Lip GYH, Greenfield S, Raftery J, Mant J, *et al*.

No. 36
A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food.
By Abubakar I, Irvine L, Aldus CF, Wyatt GM, Fordham R, Schelenz S, *et al*.

No. 37
A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs. The SANAD trial.
By Marson AG, Appleton R, Baker GA, Chadwick DW, Doughty J, Eaton B, *et al*.

No. 38
Clinical effectiveness and cost-effectiveness of different models of managing long-term oral anticoagulation therapy: a systematic review and economic modelling.
By Connock M, Stevens C, Fry-Smith A, Jowett S, Fitzmaurice D, Moore D, *et al*.

No. 39
A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder.
By Soares-Weiser K, Bravo Vergel Y, Beynon S, Dunn G, Barbieri M, Duffy S, *et al*.

No. 40
Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation.
By Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A.

No. 41
The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation.
By Burr JM, Mowatt G, Hernández R, Siddiqui MAR, Cook J, Lourenco T, *et al*.

No. 42
Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models.
By Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I.

No. 43
Contamination in trials of educational interventions.
By Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al*.

No. 44
Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers.
By Facey K, Bradbury I, Laking G, Payne E.

No. 45
The effectiveness and cost-effectiveness of carmustine implants and temozolomide for the treatment of newly diagnosed high-grade glioma: a systematic review and economic evaluation.
By Garside R, Pitt M, Anderson R, Rogers G, Dyer M, Mealing S, *et al*.

No. 46
Drug-eluting stents: a systematic review and economic evaluation.
By Hill RA, Boland A, Dickson R, Dündar Y, Haycox A, McLeod C, *et al*.

No. 47
The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model.
By Fox M, Mealing S, Anderson R, Dean J, Stein K, Price A, *et al*.

No. 48
Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study.
By Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, *et al*.

No. 49
Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial.
By Sharples L, Hughes V, Crean A, Dyer M, Buxton M, Goldsmith K, *et al*.

No. 50
Evaluation of diagnostic tests when there is no gold standard. A review of methods.
By Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM.

No. 51
Systematic reviews of the clinical effectiveness and cost-effectiveness of proton pump inhibitors in acute upper gastrointestinal bleeding.
By Leontiadis GI, Sreedharan A, Dorward S, Barton P, Delaney B, Howden CW, *et al*.

No. 52
A review and critique of modelling in prioritising and designing screening programmes.
By Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, *et al*.

Health Technology Assessment Programme

Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Deputy Director,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield,
School of Health and Related
Research

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Professor Bruce Campbell,
Consultant Vascular & General
Surgeon, Royal Devon & Exeter
Hospital

Professor Robin E Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Dr Edmund Jessop, Medical
Adviser, National Specialist,
Commissioning Advisory Group
(NSCAG), Department of
Health, London

Professor Jon Nicholl, Director,
Medical Care Research Unit,
University of Sheffield,
School of Health and
Related Research

Dr Ron Zimmern, Director,
Public Health Genetics Unit,
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield,
School of Health and Related
Research

Deputy Chair,
Dr Andrew Farmer,
University Lecturer in General
Practice, Department of
Primary Health Care,
University of Oxford

Dr Jeffrey Aronson,
Reader in Clinical
Pharmacology, Department of
Clinical Pharmacology,
Radcliffe Infirmary, Oxford

Professor Deborah Ashby,
Professor of Medical Statistics,
Department of Environmental
and Preventative Medicine,
Queen Mary University of
London

Professor Ann Bowling,
Professor of Health Services
Research, Primary Care and
Population Studies,
University College London

Professor John Cairns,
Professor of Health Economics,
Public Health Policy,
London School of Hygiene
and Tropical Medicine,
London

Professor Nicky Cullum,
Director of Centre for Evidence
Based Nursing, Department of
Health Sciences, University of
York

Professor Jon Deeks,
Professor of Health Statistics,
University of Birmingham

Professor Jenny Donovan,
Professor of Social Medicine,
Department of Social Medicine,
University of Bristol

Professor Freddie Hamdy,
Professor of Urology,
University of Sheffield

Professor Allan House,
Professor of Liaison Psychiatry,
University of Leeds

Professor Sallie Lamb, Director,
Warwick Clinical Trials Unit,
University of Warwick

Professor Stuart Logan,
Director of Health & Social
Care Research, The Peninsula
Medical School, Universities of
Exeter & Plymouth

Professor Miranda Mugford,
Professor of Health Economics,
University of East Anglia

Dr Linda Patterson,
Consultant Physician,
Department of Medicine,
Burnley General Hospital

Professor Ian Roberts,
Professor of Epidemiology &
Public Health, Intervention
Research Unit, London School
of Hygiene and Tropical
Medicine

Professor Mark Sculpher,
Professor of Health Economics,
Centre for Health Economics,
Institute for Research in the
Social Services,
University of York

Professor Kate Thomas,
Professor of Complementary
and Alternative Medicine,
University of Leeds

Professor David John Torgerson,
Director of York Trial Unit,
Department of Health Sciences,
University of York

Professor Hywel Williams,
Professor of
Dermato-Epidemiology,
University of Nottingham

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies & Screening Panel

Chair,
Dr Ron Zimmern, Director of
the Public Health Genetics Unit,
Strangeways Research
Laboratories, Cambridge

Ms Norma Armston,
Freelance Consumer Advocate,
Bolton

Professor Max Bachmann,
Professor of Health Care
Interfaces, Department of
Health Policy and Practice,
University of East Anglia

Professor Rudy Bilous
Professor of Clinical Medicine &
Consultant Physician,
The Academic Centre,
South Tees Hospitals NHS Trust

Ms Dea Birkett, Service User
Representative, London

Dr Paul Cockcroft, Consultant
Medical Microbiologist and
Clinical Director of Pathology,
Department of Clinical
Microbiology, St Mary's
Hospital, Portsmouth

Professor Adrian K Dixon,
Professor of Radiology,
University Department of
Radiology, University of
Cambridge Clinical School

Dr David Elliman, Consultant in
Community Child Health,
Islington PCT & Great Ormond
Street Hospital, London

Professor Glyn Elwyn,
Research Chair, Centre for
Health Sciences Research,
Cardiff University, Department
of General Practice, Cardiff

Professor Paul Glasziou,
Director, Centre for
Evidence-Based Practice,
University of Oxford

Dr Jennifer J Kurinczuk,
Consultant Clinical
Epidemiologist, National
Perinatal Epidemiology Unit,
Oxford

Dr Susanne M Ludgate,
Clinical Director, Medicines &
Healthcare Products Regulatory
Agency, London

Mr Stephen Pilling, Director,
Centre for Outcomes,
Research & Effectiveness,
Joint Director, National
Collaborating Centre for Mental
Health, University College
London

Mrs Una Rennard,
Service User Representative,
Oxford

Dr Phil Shackley, Senior
Lecturer in Health Economics,
Academic Vascular Unit,
University of Sheffield

Dr Margaret Somerville,
Director of Public Health
Learning, Peninsula Medical
School, University of Plymouth

Dr Graham Taylor, Scientific
Director & Senior Lecturer,
Regional DNA Laboratory, The
Leeds Teaching Hospitals

Professor Lindsay Wilson
Turnbull, Scientific Director,
Centre for MR Investigations &
YCR Professor of Radiology,
University of Hull

Professor Martin J Whittle,
Clinical Co-director, National
Co-ordinating Centre for
Women's and Childhealth

Dr Dennis Wright,
Consultant Biochemist &
Clinical Director,
The North West London
Hospitals NHS Trust,
Middlesex

Pharmaceuticals Panel

Chair,
Professor Robin Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Ms Anne Baileff, Consultant
Nurse in First Contact Care,
Southampton City Primary Care
Trust, University of
Southampton

Professor Imti Choonara,
Professor in Child Health,
Academic Division of Child
Health, University of
Nottingham

Professor John Geddes,
Professor of Epidemiological
Psychiatry, University of
Oxford

Mrs Barbara Greggains,
Non-Executive Director,
Greggains Management Ltd

Dr Bill Gutteridge, Medical
Adviser, National Specialist
Commissioning Advisory Group
(NSCAG), London

Mrs Sharon Hart,
Consultant Pharmaceutical
Adviser, Reading

Dr Jonathan Karnon, Senior
Research Fellow, Health
Economics and Decision
Science, University of Sheffield

Dr Yoon Loke, Senior Lecturer
in Clinical Pharmacology,
University of East Anglia

Ms Barbara Meredith,
Lay Member, Epsom

Dr Andrew Prentice, Senior
Lecturer and Consultant
Obstetrician & Gynaecologist,
Department of Obstetrics &
Gynaecology, University of
Cambridge

Dr Frances Rotblat, CPMP
Delegate, Medicines &
Healthcare Products Regulatory
Agency, London

Dr Martin Shelly,
General Practitioner,
Leeds

Mrs Katrina Simister, Assistant
Director New Medicines,
National Prescribing Centre,
Liverpool

Dr Richard Tiner, Medical
Director, Medical Department,
Association of the British
Pharmaceutical Industry,
London

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and
General Surgeon, Department
of Surgery, Royal Devon &
Exeter Hospital

Dr Mahmood Adil, Deputy
Regional Director of Public
Health, Department of Health,
Manchester

Dr Aileen Clarke,
Consultant in Public Health,
Public Health Resource Unit,
Oxford

Professor Matthew Cooke,
Professor of Emergency
Medicine, Warwick Emergency
Care and Rehabilitation,
University of Warwick

Mr Mark Emberton, Senior
Lecturer in Oncological
Urology, Institute of Urology,
University College Hospital

Professor Paul Gregg,
Professor of Orthopaedic
Surgical Science, Department of
General Practice and Primary
Care, South Tees Hospital NHS
Trust, Middlesbrough

Ms Maryann L Hardy,
Lecturer, Division of
Radiography, University of
Bradford

Dr Simon de Lusignan,
Senior Lecturer, Primary Care
Informatics, Department of
Community Health Sciences,
St George's Hospital Medical
School, London

Dr Peter Martin, Consultant
Neurologist, Addenbrooke's
Hospital, Cambridge

Professor Neil McIntosh,
Edward Clark Professor of Child
Life & Health, Department of
Child Life & Health, University
of Edinburgh

Professor Jim Neilson,
Professor of Obstetrics and
Gynaecology, Department of
Obstetrics and Gynaecology,
University of Liverpool

Dr John C Pounsford,
Consultant Physician,
Directorate of Medical Services,
North Bristol NHS Trust

Dr Karen Roberts, Nurse
Consultant, Queen Elizabeth
Hospital, Gateshead

Dr Vimal Sharma, Consultant
Psychiatrist/Hon. Senior
Lecturer, Mental Health
Resource Centre, Cheshire and
Wirral Partnership NHS Trust,
Wallasey

Professor Scott Weich,
Professor of Psychiatry,
Division of Health in the
Community, University of
Warwick

Disease Prevention Panel

Members

Chair,
Dr Edmund Jessop, Medical
Adviser, National Specialist
Commissioning Advisory Group
(NSCAG), London

Mrs Sheila Clark, Chief
Executive, St James's Hospital,
Portsmouth

Mr Richard Copeland,
Lead Pharmacist: Clinical
Economy/Interface,
Wansbeck General Hospital,
Northumberland

Dr Elizabeth Fellow-Smith,
Medical Director,
West London Mental Health
Trust, Middlesex

Mr Ian Flack, Director PPI
Forum Support, Council of
Ethnic Minority Voluntary
Sector Organisations,
Stratford

Dr John Jackson,
General Practitioner,
Newcastle upon Tyne

Mrs Veronica James, Chief
Officer, Horsham District Age
Concern, Horsham

Professor Mike Kelly,
Director, Centre for Public
Health Excellence,
National Institute for Health
and Clinical Excellence,
London

Professor Yi Mien Koh,
Director of Public Health and
Medical Director, London
NHS (North West London
Strategic Health Authority),
London

Ms Jeanett Martin,
Director of Clinical Leadership
& Quality, Lewisham PCT,
London

Dr Chris McCall, General
Practitioner, Dorset

Dr David Pencheon, Director,
Eastern Region Public Health
Observatory, Cambridge

Dr Ken Stein, Senior Clinical
Lecturer in Public Health,
Director, Peninsula Technology
Assessment Group,
University of Exeter,
Exeter

Dr Carol Tannahill, Director,
Glasgow Centre for Population
Health, Glasgow

Professor Margaret Thorogood,
Professor of Epidemiology,
University of Warwick,
Coventry

Dr Ewan Wilkinson,
Consultant in Public Health,
Royal Liverpool University
Hospital, Liverpool

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Expert Advisory Network

Members

Professor Douglas Altman,
Professor of Statistics in
Medicine, Centre for Statistics
in Medicine, University of
Oxford

Professor John Bond,
Director, Centre for Health
Services Research, University of
Newcastle upon Tyne, School of
Population & Health Sciences,
Newcastle upon Tyne

Professor Andrew Bradbury,
Professor of Vascular Surgery,
Solihull Hospital, Birmingham

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive,
Regulation and Improvement
Authority, Belfast

Ms Tracy Bury,
Project Manager, World
Confederation for Physical
Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and
Gynaecology and Head of the
School of Medicine,
University of Southampton

Dr Christine Clark,
Medical Writer & Consultant
Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing & Head of
Research, School of Health
Sciences, University of
Birmingham, Edgbaston,
Birmingham

Professor Barry Cookson,
Director, Laboratory of
Healthcare Associated Infection,
Health Protection Agency,
London

Dr Carl Counsell, Clinical
Senior Lecturer in Neurology,
Department of Medicine &
Therapeutics, University of
Aberdeen

Professor Howard Cuckle,
Professor of Reproductive
Epidemiology, Department of
Paediatrics, Obstetrics &
Gynaecology, University of
Leeds

Dr Katherine Darton,
Information Unit, MIND –
The Mental Health Charity,
London

Professor Carol Dezateux,
Professor of Paediatric
Epidemiology, London

Dr Keith Dodd, Consultant
Paediatrician, Derby

Mr John Dunning,
Consultant Cardiothoracic
Surgeon, Cardiothoracic
Surgical Unit, Papworth
Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon,
Gloucestershire Royal Hospital,
Gloucester

Professor Martin Eccles,
Professor of Clinical
Effectiveness, Centre for Health
Services Research, University of
Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community
Rehabilitation, Institute of
General Practice and Primary
Care, University of Sheffield

Professor Gene Feder, Professor
of Primary Care Research &
Development, Centre for Health
Sciences, Barts & The London
Queen Mary's School of
Medicine & Dentistry, London

Mr Leonard R Fenwick,
Chief Executive, Newcastle
upon Tyne Hospitals NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and
President, National Childbirth
Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine,
Department of Medicine,
University of Birmingham,
Queen Elizabeth Hospital,
Edgbaston, Birmingham

Dr Neville Goodman,
Consultant Anaesthetist,
Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor and Director of
Medical Oncology, Christie CRC
Research Centre, Christie
Hospital NHS Trust, Manchester

Professor Allen Hutchinson,
Director of Public Health &
Deputy Dean of ScHARR,
Department of Public Health,
University of Sheffield

Professor Peter Jones, Professor
of Psychiatry, University of
Cambridge, Cambridge

Professor Stan Kaye, Cancer
Research UK Professor of
Medical Oncology, Section of
Medicine, Royal Marsden
Hospital & Institute of Cancer
Research, Surrey

Dr Duncan Keeley,
General Practitioner (Dr Burch
& Ptnrs), The Health Centre,
Thame

Dr Donna Lamping,
Research Degrees Programme
Director & Reader in Psychology,
Health Services Research Unit,
London School of Hygiene and
Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor
Neurone Disease Association,
Northampton

Professor James Lindesay,
Professor of Psychiatry for the
Elderly, University of Leicester,
Leicester General Hospital

Professor Julian Little,
Professor of Human Genome
Epidemiology, Department of
Epidemiology & Community
Medicine, University of Ottawa

Professor Rajan Madhok,
Consultant in Public Health,
South Manchester Primary
Care Trust, Manchester

Professor Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Professor Alistaire McGuire,
Professor of Health Economics,
London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Andrew Mortimore, Public
Health Director, Southampton
City Primary Care Trust,
Southampton

Dr Sue Moss, Associate Director,
Cancer Screening Evaluation
Unit, Institute of Cancer
Research, Sutton

Mrs Julietta Patnick,
Director, NHS Cancer Screening
Programmes, Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry,
Royal South Hants Hospital,
Southampton

Professor Chris Price,
Visiting Professor in Clinical
Biochemistry, University of
Oxford

Professor William Rosenberg,
Professor of Hepatology and
Consultant Physician, University
of Southampton, Southampton

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh

Dr Susan Schonfield, Consultant
in Public Health, Hillingdon
PCT, Middlesex

Dr Eamonn Sheridan,
Consultant in Clinical Genetics,
Genetics Department,
St James's University Hospital,
Leeds

Professor Sarah Stewart-Brown,
Professor of Public Health,
University of Warwick,
Division of Health in the
Community Warwick Medical
School, LWMS, Coventry

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick

Dr Ross Taylor,
Senior Lecturer, Department of
General Practice and Primary
Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA –
Expert Advisory Network

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.