

Economic modelling of diagnostic and treatment pathways in NICE clinical guidelines: the MAPGuide project

Draft accepted 14 April 2013 – Not for quotation

Authorship

Joanne Lord¹, Sarah Willis², Julie Eatock³, Paul Tappenden⁴, Marta Trapero-Bertran¹, Alec Miners², Catriona Crossan¹, Maggie Westby⁵, Anastasia Anagnostou³, Simon Taylor³, Ifigeneia Mavranezouli⁶, David Wonderling⁵, Philip Alderson⁷, Francis Ruiz⁷

1. Health Economics Research Group, Brunel University.
2. Health Services Research Unit, London School of Hygiene and Tropical Medicine.
3. Information Systems, Computing and Maths, Brunel University
4. School of Health and Related Research (SchARR), University of Sheffield.
5. National Clinical Guideline Centre, Royal College of Physicians.
6. National Collaborating Centre for Mental Health, University College London.
7. National Institute for Health and Clinical Excellence, London

Corresponding author: Joanne Lord
Health Economics Research Group
Brunel University, Uxbridge UB8 3PH
joanne.lord@brunel.ac.uk

Competing interests

All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare (1) no financial support for the submitted work from anyone other than their employer; (2) no financial relationships with commercial entities that might have an interest in the submitted work; (3) no spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; and (4) no non-financial interests that may be relevant to the submitted work.

Contributions of authors

All authors contributed to the conception and design of the study or the analysis and interpretation of the data, drafting or revising the report, and final approval of the version to be published.

Joanne Lord (Reader, Health Economics) led the project and supervised development of the atrial fibrillation model.

Sarah Willis (Research Fellow, Health Economics) co-developed the prostate cancer model and analysed potential update topics.

Julie Eatock (Research Fellow, Simulation Modelling) co-developed the atrial fibrillation model and programmed the model.

Paul Tappenden (Senior Research Fellow, Health Economics and Modelling) helped design, implement and validate the prostate cancer model.

Marta Trapero-Bertran (Research Fellow, Health Economics) co-developed the atrial fibrillation model and obtained data for model.

Alec Miners (Lecturer, Health Economics) supervised development of the prostate cancer model.

Catriona Crossan (Research Assistant, Health Economics) collated potential update topics and conducted the survey of stakeholders.

Maggie Westby (Clinical Effectiveness Lead, Systematic Reviewer) contributed to collation of potential update topics and stakeholder survey, and advised on systematic reviewing in guidelines.

Anastasia Anagnostou (Research Student) contributed to the atrial fibrillation analysis.

Simon Taylor (Reader, Information Systems) advised on simulation modelling.

Ifigeneia Mavranzouli (Senior Health Economist) advised on selection of potential update topics and use of economics in NICE guidelines.

David Wonderling (Health Economics Lead) advised on selection of potential update topics and use of economics in NICE guidelines.

Philip Alderson (Associate Director, Medical Statistician and Systematic Reviewer) advised on selection of potential update topics and development of guidelines.

Francis Ruiz (Senior Advisor, Health Economics) advised on use of economics in NICE guidelines.

Acknowledgements

We would like to thank the following individuals who helped in the conduct of this research.

Expert advice for the atrial fibrillation model was provided by: Dr Derick Todd, Consultant Cardiologist, Liverpool Heart and Chest Hospital; Dr Tom Marshall, General Practitioner and Senior Lecturer in Public Health, University of Birmingham; Dr Richard Grocott-Mason, Cardiologist, Royal Brompton & Harefield Trust; and Dr Wajid Hussain, Consultant Cardiologist, Royal Brompton & Harefield Trust. Dr Marshall also assisted in the development of the atrial fibrillation model by provision of data from the THIN database. External review of the model was provided by Ben Kearns, ScHARR, University of Sheffield.

Expert advice for the prostate cancer model was provided by: Dr Chris Parker, Consultant Clinical Oncologist, The Royal Marsden; Dr. Fady Youssef, Urological Registrar, Doncaster Royal Infirmary; Dr John Graham, Lead Clinician of the NICE Prostate cancer guideline CG58 development group. UK Cancer Registry data was provided by Luke Hounslow from the South West Public Health Observatory (SWPHO).

This work was supported by the Medical Research Council [grant number G0901504].

The authors wish to acknowledge that support for Julie Eatock and Anastasia Anagnostou was provided by the EPSRC through the MATCH programme (EP/F063822/1 and EP/G012393/1).

The views expressed are those of the authors alone.

Table of contents

Executive Summary.....	i
List of abbreviations.....	viii
List of tables.....	xi
List of figures.....	xiii
Chapter 1. Background	1
Introduction to clinical guidelines.....	1
The NICE clinical guidelines programme.....	6
Methods for modelling in guidelines	11
The MAPGuide project	18
Chapter 2. Overview of methods.....	23
Selection of case studies	23
Identification of potential update topics	24
Stakeholder survey.....	24
Model development.....	24
Cost-effectiveness analysis	28
Chapter 3. Stakeholder Surveys.....	35
Introduction.....	35
Aims and objectives	35
Methods	36
Results	40
Discussion.....	45
Chapter 4. Case study 1: Full guideline model for prostate cancer.....	49
Introduction.....	49
Methods	50
Results	89
Discussion.....	100

Chapter 5. Case study 2: Full guideline model for atrial fibrillation	108
Introduction.....	108
Background.....	108
Methods	109
Results	164
Discussion.....	184
Chapter 6. Discussion.....	191
Motivation for the MAPGuide project	191
Summary of main findings	192
Implications for guideline development	201
Recommendations for research.....	205
References	207
Appendix 1. MAPGuide study protocol	225
Appendix 2. Guidelines considered as case studies	236
Appendix 3. Stakeholder survey responses	239
Appendix 4. Prostate cancer service pathway.....	251
Appendix 5. Prostate cancer data for update topics	254
Appendix 6. Atrial Fibrillation clinical pathway	259

Executive Summary

Background

NICE guidelines make recommendations for the care of people with defined diseases or conditions in the NHS in England and Wales. The recommendations are drawn up by a Guideline Development Group (GDG) after consideration of evidence of the benefits, harms and costs of various diagnostic and treatment options. As NICE guidelines are usually very broad, covering large complex pathways of care for heterogeneous groups of patients, the approach to assessing clinical and economic evidence is inevitably selective. The GDG identifies key 'clinical questions' within the scope, and evidence for these questions is identified, critiqued and summarised in a series of systematic reviews. Evidence of cost-effectiveness is drawn from relevant literature or from original evaluations conducted by guideline economists. Timelines and resources are not usually sufficient to build new models to evaluate the cost-effectiveness of every clinical question, and so specific topics for economic analysis are prioritised. There are three potential risks with this approach: 1) estimates of cost-effectiveness might not be available for some topics with potentially important resource consequences; 2) methods and assumptions used for economic analysis at different points in the pathway might be inconsistent; and 3) systemic effects and interactions between decisions might be neglected.

Another approach to economic evaluation in clinical guidelines would be to build a model of the full care pathway and to use this as a platform to evaluate changes that are being contemplated. This has some putative advantages. Though a big initial investment of time and resources, once developed the full guideline model should enable analysis of a range of cost-effectiveness questions, and provide the flexibility to address unanticipated questions, which often arise during or after guideline development. The full guideline model would also provide a common framework of methods, baseline data and assumptions, to improve the consistency of cost-effectiveness estimates. Furthermore, embedding the decision options within the context of the whole pathway provides the potential to account for systemic interactions between options. However, there are some technical and practical obstacles to the adoption of a full guideline modelling approach in the NICE clinical guidelines programme. To reflect guideline pathways, the models would most likely need to

be large and complex, and would therefore be challenging to build and to validate. It is therefore unclear whether full pathway modelling in NICE clinical guidelines is realistic, and what value it would add to the more conventional piecemeal approach to economic analysis.

Objectives

The Modelling Algorithm Pathways in Guidelines (MAPGuide) project was set up to test the feasibility and usefulness of full pathway modelling in NICE clinical guidelines. We chose to test the concept first outside of routine guideline development by building models for two existing guidelines. The guidelines chosen as case studies were both due to be reviewed by NICE to determine whether they should be updated. This offered an opportunity to identify new evidence and suggested changes to the guidelines raised by experts and stakeholders, which provided some real cost-effectiveness questions to be addressed by our models. Stakeholders for the two guidelines were surveyed to elicit their views on the relative importance of the topics as a comparison for the economic priorities identified by the models.

The aims of the project were therefore to:

- Investigate the feasibility of modelling whole service pathways from NICE clinical guidelines to estimate the cost-effectiveness of possible changes to the pathways.
- Use this approach to estimate the potential value of updating selected topics within the guidelines, as estimated by the likelihood of a change in recommendation and the potential impact of any such change on expected net benefits.
- Compare the economic priorities for updates obtained from formal modelling with those elicited during the routine NICE guideline review process.

Methods

Two published NICE guidelines were selected as case studies: prostate cancer and atrial fibrillation. These guidelines were chosen from the 17 due to be reviewed by NICE within our timelines (between January and September 2011). The criteria for selecting the case studies were: a) that the published guideline contained a relatively well-formulated pathway suitable for modelling; b) that they addressed important topics that were likely to be updated; c) that there was thought to be uncertainty or controversy over which topics should be updated; and d) the guidelines would address different patient groups or disease areas, and hence would present different challenges for the modellers.

The project consisted of three streams of work, led by different teams of researchers: the first team (CC, MW) identified the potential update topics and conducted and analysed the surveys of stakeholders; the second team (SW, PT, AM) developed the prostate cancer model and used it to evaluate possible update topics for this guideline; and the third team (MTB, JE, AA and JL) developed the atrial fibrillation model and evaluated possible update topics for this disease area. The modelling teams did not influence the choice of topics and did not see their list of topics until the design of the base case model had been finalised. They were also not shown the results of the stakeholder surveys until after the base case models had been implemented. The two modelling teams met regularly to discuss progress and difficulties encountered, and to agree the general principles to be followed.

The potential update topics were elicited from the review decision documents published on the NICE website. One researcher (CC) read the documents and collated a list of suggested topics with advice from a second researcher who has extensive experience in systematic reviewing and guideline development (MW). A shortlist of topics for each guideline was agreed by members of the research team who were not involved in developing the models. Information about the selected topics and relevant new evidence cited in the review documents was collated. Surveys were then conducted with registered stakeholders for the two guidelines to elicit their opinions about the importance of including the selected topics in a future update of the guidelines. Participants were presented with a short summary of the shortlisted topics and then asked to rate them in terms of importance (using a five point Likert scale) and to rank them in order of priority for inclusion in a future update. Results were summarised using simple descriptive statistics and graphs.

The modelling teams began with background reading to familiarise themselves with the guideline and current issues in the field. The boundaries for the modelling exercises were defined by the scopes for the original guidelines. The model designs consisted of two main elements: a 'service pathway model', which specifies the expected care that patients with defined characteristics would receive according to the current guideline recommendations; and a 'disease process model', which specifies how patients' health status or risk of events is expected to change conditional on their characteristics and treatment. The pathway models were developed following detailed examination of the guideline documents, and then checked with clinical experts. Published NICE technology appraisal guidance within the

scope of the guidelines was added to the care pathways. The disease process models were developed following review of other published models, descriptions of epidemiology and discussions with clinical experts. The implemented models were designed to estimate the mean discounted healthcare costs and outcomes (QALYs) over a lifetime horizon for a representative heterogeneous cohort of patients treated according to the defined pathway, and following the NICE Reference Case for economic evaluations. The models were implemented as individual-level Discrete Event Simulations (DES) programmed using a specialist software package (Simul8).

The starting characteristics for patients entering the models were taken from samples of individual data from a general practice database for atrial fibrillation and a disease register for prostate cancer. It was beyond the scope of this project to conduct systematic reviews to inform all model parameters. Information was therefore identified from the original guideline, supplemented with new evidence from rapid reviews of the literature and discussions with clinical experts. The impact of uncertainty over model parameters was considered through probabilistic sensitivity analysis. The models were each evaluated over two nested levels of iteration: an internal level to reflect variability between individual patients (first order uncertainty); and an external level to represent uncertainty over population parameters (second order uncertainty). The modelling teams checked for errors and inconsistencies during model development. The models were also verified to test for correct programming and validated to ensure consistency with expected results – for example, that survival times and levels of service use were realistic.

The modelling teams then attempted to estimate the cost-effectiveness of questions related to their list of possible update topics. This involved running the models for the base case pathways, and then for a series of strategies reflecting possible changes to the pathways. Additional data required for the alternative strategies were obtained from the original guideline, new evidence identified by stakeholders or experts, or by rapid literature searches. It is important to note that the economic analyses presented in this report were not based on full evidence reviews and have not been informed by an expert guideline group. Our focus was to assess the modelling methods and their usefulness in the context of identifying priorities for inclusion in a guideline update. The results should not be used to inform clinical practice.

The optimal strategy within each topic was identified by incremental cost-effectiveness analysis, or equivalently by calculation of incremental net benefits. Results are reported using a primary cost-effectiveness threshold of £20,000 per QALY gained. The relative 'economic priority' for which topics should be included in an update was determined on the basis of two pieces of information: i) the probability that the currently-recommended base case option is not the optimal strategy within that topic and ii) the magnitude of the potential gain in net benefit (difference between the optimum and base case strategies). These statistics were used by the modelling teams to rank the topics in order of importance for inclusion in an update. This ranking was then compared with that obtained from the survey of guideline stakeholders.

Results

Both modelling teams succeeded in building a discrete event simulation representing the guideline pathway and disease process. Development took nearly 24 months, and required more than one whole time equivalent years of analyst time for each full guideline model. The scope of both models was very broad, and included the large majority of the guideline pathways and recommendations, though with some exceptions.

Nine potential update topics were identified from the NICE review documents for the prostate cancer guideline, and eight topics for the atrial fibrillation guideline. Survey responses were received from 27 out of 239 registered stakeholders for prostate cancer. Two topics were rated as 'very important' for inclusion in an update of the guideline by the majority of respondents: active surveillance in previously unscreened 'low risk' men; and effective techniques for performing radical prostatectomy. For atrial fibrillation, 32 out of 182 stakeholders responded to the survey, and two topics were rated as 'very important' by the majority of these respondents: new oral anticoagulants; and stratification tools to assess bleeding risk before antithrombotic medication.

Cost-effectiveness analysis was conducted with the prostate cancer model for six of the nine potential update topics. Two of the remaining topics were not modelled because of an absence of necessary information, and there was no need for cost-effectiveness analysis of the final topic due to likely dominance of one of the options. The model indicated that for three topics in particular, there was a high potential for increased net benefit: brachytherapy with external-beam radiotherapy for localised or locally advanced disease;

pelvic radiotherapy with adjuvant hormone therapy for localised disease; and continuous versus intermittent hormone therapy for metastatic prostate cancer.

For atrial fibrillation, cost-effectiveness estimates were obtained for five out of eight topics. Two topics were not evaluated as they related to areas excluded from the scope of the model, and one topic was omitted due to time constraints. The results for the five modelled topics suggested that stroke and bleeding risk thresholds for oral anticoagulation, and the comparison of rate and rhythm control strategies should be treated as priorities for inclusion in an update; as there was a fair probability that the current guideline recommendations were not optimal and a high potential for health gain and cost savings. The estimated economic priorities for update topics differed from those elicited from stakeholders.

Conclusions

This project has demonstrated that models can be developed to cover the large part of complicated clinical guideline pathways, and that these models can be used to evaluate a range of cost-effectiveness questions across the pathway. This approach has the potential to produce consistent estimates of cost-effectiveness that account for systemic effects of placing decision options within a broad pathway of care. Implementation of the models was facilitated by the use of discrete event simulation, which can provide a relatively compact representation of the flow of a heterogeneous cohort of patients through a very complex care pathway.

However, there are some barriers to the routine adoption of this approach in the NICE clinical guidelines programme. The most obvious is that it is unlikely to be possible within current timelines and health economic resources, because of the demands of developing such large models and the need for specialist DES expertise. It is also unlikely that this approach will work for all clinical guidelines, due to data limitations. Nevertheless, we believe that the approach does show sufficient promise to warrant further investigation.

Learning from this project should enable faster development of full guideline models in the future. The two case studies provide a set of methods and terminology that could be adapted for other applications. We also identify some considerations for future development including: the need for clarity about the boundaries of the model; the need

for clarity over whether the model pathway is meant to reflect recommended or current practice; and the importance of conceptual modelling of both the service pathway and of the disease process

We recommend five priorities for further research:

1. The case study models have been made available to the NCC teams developing updates of the guidelines. Research should be conducted to seek the opinions of the members of the NCC technical teams, the GDGs, stakeholders and the NICE guidelines team to determine whether they made use of the models, and if so whether the results were useful.
2. Further development of the case study models to assess the existence and magnitude of possible interactions between changes to different parts of the care pathways.
3. Extension of the case study models to estimate budget and health impacts across a whole patient population (rather than for single incident cohort).
4. To apply the full guideline modelling approach to a new NICE guideline.
5. Further development of methods:
 - a. Methods for eliciting expert opinion and reaching consensus about the structure of disease process and service pathway models to inform guideline development and economic modelling;
 - b. Methods for robust model calibration to infer missing or unobservable parameters in complex decision models;
 - c. Development of standardised software templates or methods of presentation to help guideline economists to develop flexible and accessible guideline models in consultation with other methodologists, GDGs and stakeholders;
 - d. Methods to test the internal and external validity of full guideline models.

Word count: 2522

List of abbreviations

AAD	Anti-arrhythmic drugs
ACE inhibitors	Angiotensin-converting-enzyme inhibitor
AF	Atrial Fibrillation
AFFIRM	Atrial Fibrillation Follow-up Investigation of Rhythm Management trial
AGREE	Appraisal of Guidelines for Research and Evaluation.
ARB	Angiotensin Receptor Blockers
BB	Beta Blocker
BMI	Body Mass Index
bpm	Beats per minute
CAF	Chronic atrial fibrillation
CCA	Cost Consequence Analysis
CEA	Cost Effectiveness Analysis
CHD	Coronary Heart Disease
CHF	Congestive Heart Failure
CRD	Center for Reviews and Dissemination, University of York
CT	Computerised Tomography
CG	Clinical Guideline
DALY	Disability Adjusted Life Year
DES	Discrete Event Simulation
DTI	Direct Thrombin Inhibitor
ECG	Electrocardiography
ECV	Electrical cardioversion
EQ-5D	EuroQol 5D outcome measurement instrument
ESC	European Society of Cardiology
ER ECG	Event recorder electrocardiography
FAQ	Frequently Asked Questions
GDG	Guideline Development Group
G-I-N	Guidelines International Network.
GRADE	Grading of Recommendations Assessment Development and Evaluation.
HDR	High-dose-rate brachytherapy
HTA	Health Technology Assessment

HTAi	Health Technology Assessment International
ICER	Incremental Cost Effectiveness Ratio
ICG	Internal Clinical Guidelines
IGRT	Image Guided Radiation Therapy
IMRT	Intensity Modulated Radiation Therapy
INB	Incremental Net Benefit
INR	International Normalized Ratio
IOM	Institute of Medicine
LDR	Low-dose rate brachytherapy
LHRH	Luteinizing hormone-releasing hormone
LYG	Life Year Gained
LVD	Left Ventricular Dysfunction
MAPGuide	Modelling Algorithm Pathways in Guidelines
MI	Myocardial Infarction
MRC	Medical Research Council
NB	Net Benefit
NCC	National Collaborating Centre
NCC-C	National Collaborating Centre for Cancer
NCCMH	National Collaborating Centre for Mental Health
NCC-WCH	National Collaborating Centre for Women's and Children's Health
NCGC	National Clinical Guideline Centre
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence.
NIHR	National Institute for Health Research
NSAID	Non steroidal anti inflammatory drug
OAC	Oral anticoagulant
ONS	Office for National Statistics
OXVASC	Oxford Vascular Study
PCV	Pharmacological cardioversion
PET	Positron emission tomography
PSA	Prostate Specific Antigen
PSS	Personal Social Services
QALY	Quality Adjusted Life Year

QRG	Quick Reference Guide
RCT	Randomised Controlled Trial
RealiseAF	Real-life global survey evaluating patients with Atrial Fibrillation
RLCA	Rate limiting calcium antagonist
SHD	Structural heart disease
SR	Sinus rhythm
TA	Technology Appraisal
TE	Thromboembolism
TIA	Transient Ischaemic Attack
THIN	The Health Improvement Network
TOE	Transesophageal Echocardiogram
TTE	Transthoracic Echocardiogram

List of tables

Table 1. Responses to prostate cancer survey	42
Table 2. Responses to atrial fibrillation Survey.....	42
Table 3 . Top three prostate cancer topics	43
Table 4. Top three atrial fibrillation topics	43
Table 5. D’Amico risk classification ⁷⁰	53
Table 6: Radical treatment efficacy and adverse event parameters.....	73
Table 7: Palliative treatment efficacy parameters	74
Table 8: Utility data.....	75
Table 9: Resource use parameters	76
Table 10. Modelling scenarios in Hummel <i>et al.</i> 2010 ¹⁰²	85
Table 11. Summary of amendments to evaluate update topics	87
Table 12. Intermediate results (per 1,000 men referred)	89
Table 13. Health outcomes by base case model segment (per 1,000 men referred)	90
Table 14. Costs from the base case model (per 1,000 men referred).....	90
Table 15. Summary of mean costs and outcomes for all options tested (per 1,000 men referred).....	92
Table 16. Incremental cost-effectiveness results - Topic A (per 1,000 men referred).....	93
Table 17. Incremental cost-effectiveness results - Topic B (per 1,000 men referred).....	95
Table 18. Incremental cost-effectiveness results - Topic C (per 1,000 men referred)	96
Table 19. Incremental cost-effectiveness results - Topic F (per 1,000 men referred)	97
Table 20. Incremental cost-effectiveness results - Topic H (per 1,000 men referred).....	98
Table 21. Summary of net benefit results (per 1,000 men referred)	100
Table 22. Risk factors drawn from THIN data	130
Table 23 Choice of antithrombotic drugs by risk factors: age 65-74.....	134
Table 24. Drugs currently included in the model	135
Table 25. Rates of thromboembolism from the Swedish AF cohort study ¹⁴⁹	143
Table 26. Rates of bleeding from the Swedish AF cohort study ¹⁴⁹	143
Table 27. Cardioversion success rates	144
Table 28. AF recurrence after cardioversion	144
Table 29. Treatment effects of anti-thrombotic drugs ¹³⁶	145

Table 30. Treatment effects of anti-arrhythmic drugs ¹⁷⁰	145
Table 31. Treatment effects of rate control drugs ¹⁷¹⁻¹⁷⁴	146
Table 32. General population utility by age (no other heart problems) ¹⁷⁸	147
Table 33. Utility multipliers for AF health states, RealiseAF ¹⁷⁹	147
Table 34. Utility multipliers for adverse events, MEPS ¹⁸⁰	148
Table 35. Costs for consultations and tests on the pathway	148
Table 36. Daily drug costs (NHS Net Price, BNF ¹⁰⁶)	149
Table 37. Cost of adverse events included in model	150
Table 38. Topics for atrial fibrillation model	153
Table 39. Results for 10,000 patients: base case model, deterministic	166
Table 40. Base case model results (per 1,000 incident cases)	172
Table 41. Model results: PCV drugs for patients without structural heart disease (per 1,000 patients)	174
Table 42. Model results: rate vs. rhythm control for persistent AF (per 1,000 incident cases)	175
Table 43. Model results: CHA ₂ DS ₂ -VASc thresholds for warfarin (per 1,000 incident cases)	176
Table 44. Model results: CHADS ₂ thresholds for warfarin (per 1,000 incident cases)	176
Table 45. Model results: HAS-BLED thresholds for warfarin (per 1,000 incident cases)	178
Table 46. Model results: comparison of oral anticoagulant drugs (per 1,000 incident cases)	178
Table 47. Model results: HAS-BLED and CHA ₂ DS ₂ -VASc thresholds for warfarin	180
Table 48. Model results: HAS-BLED and CHADS ₂ thresholds for warfarin	181
Table 49. Model results: HAS-BLED and CHADS ₂ thresholds for dabigatran	182
Table 50. Model results: HAS-BLED and CHADS ₂ thresholds for rivaroxaban	183

List of figures

Figure 1 Summary of the service pathway based on guideline recommendations	55
Figure 2. Disease process model.....	56
Figure 3. Design-oriented model: pre-clinical and diagnostic sections.....	60
Figure 4. Design-oriented model: treatment pathways for diagnosed disease	61
Figure 5. Time segments (numbered) used to calculate overall survival	68
Figure 6. Observed vs. calibrated frequencies of disease events.....	71
Figure 7. Model stability according to size of simulated cohort.....	79
Figure 8. Location of topics (numbered) in the clinical pathway	80
Figure 9. RCT evidence network of surgical techniques for localised prostate cancer	81
Figure 10. Cost-effectiveness acceptability curves - Topic A.....	94
Figure 11. Cost-effectiveness acceptability curves - Topic B.....	95
Figure 12. Cost-effectiveness acceptability curve - Topic C.....	97
Figure 13. Cost-effectiveness acceptability curve for Topic F	98
Figure 14. Cost-effectiveness acceptability curves - Topic H.....	99
Figure 15. Overview of AF service pathway model.	114
Figure 16. Overview of AF service pathway model	117
Figure 17. Decision tree showing detail of diagnostic pathway	118
Figure 18. Atrial fibrillation disease process model.....	125
Figure 19. Model of AF progression and control	127
Figure 20. Illustration of ongoing management process	136
Figure 21. Illustration of the sampling process for the AF model	138
Figure 22. Age at diagnosis: 10,000 patients from THIN AF cohort.....	167
Figure 23. Simulated life years: 10,000 patients from THIN AF cohort.....	167
Figure 24. Simulated QALYs: 10,000 patients from THIN AF cohort	167
Figure 25. Simulated medication costs: 10,000 patients from THIN AF cohort	168
Figure 26. Simulated healthcare costs: 10,000 patients from THIN AF cohort	168
Figure 27. Cumulative net benefit per patient with increasing sample size	169
Figure 28. Cumulative net benefit by probabilistic iteration (1,000 patients per iteration)	173
Figure 29. Costs and effects for CHADS ₂ and CHA ₂ DS ₂ -VASc thresholds for warfarin	177

Chapter 1. Background

Introduction to clinical guidelines

Evolution of evidence based guidelines

Guidelines on clinical practice have been developed by professional bodies in many countries for many years now. Initially based on informal consensus and expert opinion, the influence of evidence based medicine has led to the adoption of more formalised methods of development. In 1992, the US Institute of Medicine (IOM) defined guidelines as:

“... systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”³

In their 2011 statement, the IOM revised their definition to:

“Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.”⁴

In addition to the use of systematic reviews, they define criteria for guidelines ‘we can trust’, including transparency, composition of the guideline development group, and external review.

Several international collaborations have been established to further the use of robust processes and methods for guideline development. For example, the Appraisal of Guidelines for Research and Evaluation (AGREE) Enterprise was founded to develop and promote a critical appraisal checklist to evaluate processes of guideline development and quality of reporting (www.agreetrust.org). The Guidelines International Network (G-I-N) was founded in 2002 to provide a network for guideline developers and users, to help reduce duplication of effort and to promote best practice in guidelines (www.g-i-n.net).

The role of cost-effectiveness in guidelines

Although there is now broad agreement over the need to base clinical guidelines on formalised methods of evidence review and synthesis, the role of cost-effectiveness in guidelines is much more controversial. In 1992, the influential IOM committee debated this question.³ They concluded that developers of clinical practice guidelines “need not” use

economic criteria in drawing up recommendations on what constitutes appropriate care: not because costs should be or can be avoided, but because the committee could not agree that guideline developers are necessarily the right people to be making these judgments. Instead, they put forward the “modest proposal” that guideline developers should present information about the costs and health implications of alternative interventions to help practitioners, patients and policymakers who face resource constraints to evaluate the options. The 2011 IOM committee also discussed this issue, but chose not to comment on the role of costs or cost-effectiveness in guideline decision-making. Similarly, the G-I-N recently advised that recommendations should be based on scientific evidence of benefits, harms and ‘if possible’ costs, but did not make any more explicit statement about whether or how this information ought to be considered.⁵

A dissenting member of the 1992 IOM committee and witness to the 2011 committee, David Eddy has been a prominent advocate for the explicit consideration of costs alongside health outcomes in guidelines. He argues:

*“...health interventions are not free, people are not infinitely rich, and the budgets of programs are limited. For every dollar’s worth of health care that is consumed, a dollar will be paid. Furthermore, the costs will be paid by present and future patients”.*⁶

This argument was taken up by Alan Williams^{7;8}, who noted that to optimise outcomes across a population, guideline developers must take account of the sacrifices imposed on other current or future patients when scarce healthcare resources are devoted to a subset of patients who are the concern of a particular guideline. This requires an appreciation of the relative costs and health effects of alternative treatment options for the defined subgroups of patients; and an understanding of what health benefits could be obtained by using resources in other ways (opportunity costs). The methods of economic evaluation are designed to assist decision-makers in making such comparisons.⁹

In practice, explicit consideration of cost or cost-effectiveness in guidelines is unusual.¹⁰ A search of the National Guideline Clearinghouse online database found that of 1616 guidelines published between 2000 and 2005, only 369 (23%) included any formal cost analysis.

Assessing cost effectiveness: Profiles vs. models

Even when it is accepted that cost-effectiveness ought to influence guideline recommendations, there is still controversy over how this should be done. Eddy defined two broad approaches to designing practice policies.⁶ In what he called the ‘implicit’ approach, experts are asked to weigh up pertinent information in their heads, deliberate and reach a collective decision. This consensus process is, he argues, satisfactory for many types of decision problem, but it is inadequate when there are complex and uncertain trade-offs between costs, benefits and harms. For this type of question, Eddy proposed an ‘explicit’ approach to decision-making, characterised by *“an explicit and systematic analysis of evidence, estimation of outcomes, calculation of costs, and assessment of preferences”*. This latter approach includes formalised methods of clinical decision analytic modelling and health economic evaluation.¹¹

A related distinction between ‘profiles’ and ‘models’ has permeated discussions about how to take account of economic considerations in clinical guidelines.¹² In the profile approach, various measures of health effect derived from clinical studies are summarised alongside estimated costs. The guideline development group then discuss, interpret and weigh up this information qualitatively. In the alternative modelling approach, there is a formal processing of information to produce a quantitative summary of the expected costs and health consequences of the available options. The summary is often in the form of an Incremental Cost Effectiveness Ratio (ICER), such as the additional cost per Quality Adjusted Life Year (QALY) gained, which when compared against a benchmark cost-effectiveness threshold provides an indication of the ‘right decision’ (if not a definitive decision rule).¹³

Eccles and Mason argued against modelling in clinical guidelines, based on their experience working on the Department of Health funded, North of England guidelines in the 1990’s.¹² This view was influential in early approaches to guidelines commissioned by the National Institute for Health and Clinical Excellence (NICE), since when NICE was established in 1999 it effectively inherited the Department of Health guidelines programme. The example that Eccles and Mason used to criticise the modelling approach is the North of England guideline on anticoagulation to prevent stroke in patients with atrial fibrillation (which, coincidentally, is a question that we address in our case study in Chapter 5). The North of England guideline used a Markov decision model to estimate the cost and QALY impact of treatment

with warfarin for various groups of patients.¹⁴ The criticisms levelled at this approach by Eccles and Mason included technical limitations of the model, particularly relating to the (as it now seems) quite primitive deterministic analysis of uncertainty. However, they also argued that the use of a model detracted from the quality of interaction of the guideline group with the evidence:

*“Once the clinical problem had been scoped there was little remaining role for the group and they were not called upon to discuss the evidence or the implications of the model.”*¹²

Eccles and Mason acknowledged that there are situations where simple modelling exercises are necessary and useful to the decision-making process. But they argued that the touchstone for such exercises is ‘parsimony’, to ensure that guideline developers and users can understand and, if necessary, replicate the results. In general, they advocated the use of a simple balance sheet to present disaggregated health effects and costs for the consideration of the guideline group.

More recently, the international Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group has also promoted a profile approach as a ‘simple but powerful’ way of presenting the advantages and disadvantages of alternative management options to a guideline panel.^{15;16} The GRADE system provides a formalised process for identifying, appraising and summarising evidence relating to important outcomes, which may include differences in resource use where relevant. GRADE advise that guideline panels may ‘legitimately ignore’ information on resource use, but that if a panel chooses to consider this information, they should first assess the quality of the underlying evidence and its applicability to their particular decision problem. Although GRADE note that formal economic modelling results can help to inform judgments about the balance between positive and negative outcomes, they highlight the downsides of modelling. In particular, they state that modelling reduces transparency and that it is susceptible to bias and uncertainty arising from the many assumptions that are required and the poor quality data that are often available to support a model.

This rather negative attitude to explicit economic evaluation and modelling in the field of clinical guidelines contrasts with the predominant view in Health Technology Assessment (HTA).¹⁷ This may be because cost containment was often seen as an important motivator

for the development of HTA, or possibly because of the influence of Archie Cochrane and his reflections on effectiveness *and* efficiency.¹⁸ But also, whereas the objective of guidelines is usually defined as informing clinical decision-making and optimising patient outcomes, the objective of HTA is more clearly directed at informing policy-making and optimising population outcomes, which makes the trade-offs between alternative uses of scarce resources are more apparent. For example, Health Technology Assessment International (HTAi) defines HTA as:

“... a multidisciplinary field that addresses the health impacts of technology, considering its specific healthcare context as well as available alternatives. Contextual factors addressed by HTA include economic, organizational, social, and ethical impacts. The scope and methods of HTA may be adapted to respond to the policy needs of a particular health system.” <http://www.htai.org>

Of course, beliefs about the role of economic considerations and the use of modelling do also vary between HTA agencies and practitioners. For example, although many European agencies have been willing to bring in explicit consideration of cost-effectiveness and the use of modelling, the criteria and methods used by these agencies differ.¹⁹ Further, whilst the public and policy-makers in the United States are generally unreceptive to cost-effectiveness or modelling, it has been argued that *“US health policymakers in the private and public sectors continue in a quieter fashion to develop strategies to use evidence of comparative value.”*²⁰

Opinions on which economic evaluation methods to use in HTA do also differ amongst health economists. Some have significant concerns over technical aspects of modelling, and over the validity of the summary measures of the ICER and the QALY. For example, the use of Cost Consequence Analysis (CCA) - akin to a profile or balance sheet approach – has been proposed as a means of bringing economic evaluation more into line with society’s values.²¹ However, the predominant view among health economists active in the field of HTA is that modelling is an ‘unavoidable fact of life’, and has the clear advantage of providing an explicit and reproducible summary of the balance of benefits, harms and cost.²² While there may be legitimate concerns about the potential for inappropriate use of data, and problems with the transparency and validity of models²³, steps can be taken to minimise these dangers.²⁴⁻

There has also been discussion about the use of systematic reviews to identify and summarise published 'economic evidence' to put before decision-makers. This is standard practice in both HTA and guidelines, but it has been argued that this is a largely futile exercise, as estimates of cost or cost-effectiveness obtained in one context are rarely transferable to another.²⁷ Modelling provides a more satisfactory method for synthesising clinical and economic evidence to provide a coherent aid to decision making. These arguments might equally be applied to guideline development.

The NICE clinical guidelines programme

Purpose and scope of NICE guidelines

The development of clinical guidelines is a core function of NICE:

*"Guidance from the Institute will include guidelines for the management of certain diseases or conditions and, where appropriate, it will cover all aspects of the management of that condition – from prevention to self-care through primary care, secondary care and more specialist services."*²⁸

Between May 2001 and July 2012, NICE published 153 Clinical Guidelines, including 8 inherited guidelines commissioned by the Department of Health, and 26 updates of previously published NICE guidelines. A further 56 guidelines are currently in development.

There are NICE guidelines for diverse patient groups and conditions, including topics in mental health, women and children's health, cancer, and acute and chronic disease. Each guideline encompasses a wide range of management options for the defined patient group, including aspects of disease prevention, case identification, assessment and diagnosis, treatment, monitoring, on-going care and self-management. Although NICE has now introduced a 'short guidelines' programme, which develops guidelines with a narrower focus, most NICE guidelines are still very broad in scope, and make a large number of recommendations to the NHS. Though compliance with NICE guidelines is not compulsory, and no special funding is available to support their implementation, they are used to set standards for NHS organisations and professionals, and can have a major impact on patient care.²⁹

Process for development of NICE guidelines

NICE currently commissions guidelines from four National Collaborating Centres (NCCs): see Box 1. In addition, an Internal Clinical Guidelines (ICG) team at NICE develops the short guidelines. The main functions of the NCC and ICG teams are to convene and to provide secretariat and technical support functions to the Guideline Development Groups (GDGs).

Box 1. National Collaborating Centres

- **National Clinical Guideline Centre (NCGC)**
Hosted by the Royal College of Physicians, in partnership with the Royal College of General Practitioners, Royal College of Nursing, Royal College of Physicians and Royal College of Surgeons of England.
- **National Collaborating Centre for Cancer (NCC-C)**
Hosted by Velindre NHS Trust in Cardiff in partnership with Cardiff University and other organisations.
- **National Collaborating Centre for Women's and Children's Health (NCC-WCH)**
Hosted in a partnership led by the Royal College of Obstetricians and Gynaecologists, which includes the Royal College of Paediatrics and Child Health, Royal College of Midwives, Royal College of Nursing and a range of other stakeholders.
- **National Collaborating Centre for Mental Health (NCCMH)**
Partnership between the Royal College of Psychiatrists and the British Psychological Society.

The development process for NICE guidelines is outlined in Box 2. After referral of the topic from the Department of Health, a scope is prepared by the NCC and after consultation with stakeholders, finalised and agreed by NICE. This document defines the populations, healthcare settings and types of interventions to be included or excluded, and sets the boundaries for the work of the GDG.

The GDG is the independent advisory committee who develop the guidance. They meet over a period of 12 to 8 months, usually monthly. Unlike NICE's Technology Appraisals Committees, a GDG is specially convened for each guideline. The composition of the GDG is tailored to the guideline topic. In addition to healthcare professionals, patient or public representatives, and sometimes healthcare managers or commissioners, the GDG includes members of the NCC technical team who are responsible for conducting the evidence reviews and any related analyses. The technical team includes individuals with skills in project management, information science, systematic reviewing and health economics. The GDG has the following key functions: to define specific review questions within the guideline scope; to consider the clinical and economic evidence related to these questions; to use

expert consensus if evidence is poor or lacking; to formulate guideline recommendations; and to respond to comments from the stakeholders.

The whole process of guideline development, from referral to publication, takes 18-24 months for standard guidelines and 9-11 months for short guidelines.

Box 2. NICE guideline development process ³⁰

1. **Guideline topic referred to NICE by Department of Health**
2. **Stakeholders register interest**
National organisations representing patients and carers, and health professionals can register as stakeholders. Stakeholders are consulted throughout the guideline development process.
3. **Scope prepared**
NCC prepares the scope, setting out what the guideline will and will not cover. Following consultation with stakeholders, the scope is agreed and signed off by NICE.
4. **Guideline development group established**
Includes health professionals, representatives of patient and carer groups and technical experts.
5. **Draft guideline produced**
The GDG assesses the available evidence and makes recommendations.
6. **Consultation on the draft guideline**
Public consultation period for registered stakeholders to comment on the draft guideline.
7. **Final guideline produced**
GDG finalises the recommendations; the collaborating centre produces the final guideline.
8. **Guidance issued**
NICE formally approves the final guideline and issues its guidance to the NHS.

Around the time of publication, NICE produces implementation support tools to encourage uptake of the guideline. These include a costing tool, which identifies any significant resource impacts of recommendations and estimates the budget impact for NHS bodies, to help them in planning for implementation.

All NICE guidelines are reviewed periodically to check whether they need updating. NICE conducts a formal review of the need to update a guideline three years after publication. This involves consultation with the original GDG, collection of intelligence and focused literature searches. A draft review decision is published for consultation with stakeholders, and then finalised by NICE. This may result in a decision to update the guideline in part or in whole, not to update the guideline, to transfer it to a 'static' list, or to withdraw it. The review decision is published by NICE, and includes a summary of new evidence and topics that need to be updated, as well as full list of stakeholder comments and responses from NICE.

Methods for development of NICE guidelines

Methods for the development of NICE guidelines are specified in the Guidelines Manual.³¹

There are four key steps to assembling and interpreting the evidence base to support GDG decision-making (see Box 3).

Box 3. NICE guideline development methods³¹

1. **Formulate the review questions**
 - Structure review questions
 - Use patient experiences to inform the review questions
 - Agree the review protocols and finalise the economic plan
2. **Identify the evidence**
 - Develop search strategy for each review question
 - Search relevant databases
 - Ensure sensitivity and specificity of search
 - Consider stakeholder submissions of evidence, if applicable
3. **Review the evidence**
 - Select relevant studies
 - Assess quality of selected studies for clinical and cost effectiveness
 - Conduct new economic evaluations on selected topics
 - Update existing NICE guidance (if identified in the scope)
 - Summarise evidence and present results
4. **Develop guideline recommendations**
 - Interpret the evidence to make recommendations
 - Formulate recommendations, paying particular attention to wording
 - Identify key priorities for implementation
 - Formulate research recommendations

Cost-effectiveness in NICE guidelines

NICE has always had a clear remit to consider both clinical effectiveness and cost-effectiveness in its guidelines.²⁸ The decision-making principles employed by NICE are outlined in its Social Value Judgements paper.³² This emphasises the importance but also the boundaries of cost-effectiveness in NICE guidance.

“Those developing clinical guidelines, technology appraisals or public health guidance must take into account the relative costs and benefits of interventions (their ‘cost effectiveness’) when deciding whether or not to recommend them.”

“Decisions about whether to recommend interventions should not be based on evidence of their relative costs and benefits alone. NICE must consider other factors when developing its guidance, including the need to distribute health resources in the fairest way within society as a whole.”

To help guideline groups to take account of cost-effectiveness, a health economist is employed as part of the technical team for all NICE guidelines. It has been argued that guideline economists are too isolated within GDGs, where the majority of members are clinicians or patient representatives with a special interest in the guideline topic and may therefore be reluctant to rule against clinically effective interventions on the grounds of cost-effectiveness.³³ This claim has however been refuted by NICE and the NCCs.³⁴⁻³⁶

The methods used by guideline economists have evolved over the twelve years that NICE has been developing guidelines. As mentioned above, economists were initially discouraged from developing their own models. This has changed, and the Guidelines Manual³¹ now broadly recommends a combination of profile and modelling approaches: with models being developed to address selected questions in each guideline, and reliance on summaries of published economic evidence or GDG judgment alone for other questions. Most guidelines now include at least one original model-based economic analysis.³⁷

In addition to providing general advice to the GDG on economic issues, the guideline economist is expected to review published economic evaluations, prioritise questions for further economic analysis, and conduct *de novo* economic evaluations for selected questions. Early in the guideline development process, the economist, in discussion with the rest of the technical team and GDG, prepares an 'economic plan' that identifies the initial priorities for further economic analysis and the proposed methods for addressing these questions. The criteria for judging the value of a new economic analysis are: the overall 'importance' of the recommendation, which depends on the number of patients affected and the costs and health impacts per patient; current uncertainty over cost effectiveness; and likelihood that further economic analysis will clarify this uncertainty.

This is not always straightforward, and economic plans can and do change during guideline development.³⁷ The Guidelines Manual advises on how new economic analyses should be conducted and reported.³¹ Analyses are expected to follow the same 'Reference Case' as NICE technology appraisals.² This specifies, for example, how to measure health effects (in QALYs), the perspective to be used for costing (NHS and local authority funded personal social services services) and the rates for discounting costs and QALYs (3.5%). The Guidelines Manual also defines some general principles for modelling in guidelines (see Box 4). NICE has adopted the GRADE framework for assessing the quality of clinical evidence

within its guidelines programme³⁸, and NICE has developed a similar framework¹⁵ for reviewing and presenting cost-effectiveness estimates from published studies or new models.

Box 4. Principles for economic evaluation in NICE guidelines

- The economist should carry out the analysis in collaboration with the rest of the GDG.
- Economic analyses should be explicitly based on the guideline's review questions.
- An economic analysis should be underpinned by the best-quality clinical evidence, based upon and consistent with that identified in addressing the guideline's review question.
- The structure of the model should be discussed and agreed with the GDG
- All cost-effectiveness analyses should be validated.
- There should be the highest level of transparency in reporting methods and results.
- Incremental analysis should be used when comparing mutually exclusive options.
- Considerations of potential bias and limitations should be discussed by the GDG.
- Sensitivity analysis should be used to explore the impact of potential sources of bias and uncertainty. Probabilistic sensitivity analysis is the preferred method for taking account of uncertainty arising from imprecision in model parameters

Methods for modelling in guidelines

Critique of the current NICE approach

The selective approach to economic modelling currently used in NICE guidelines is pragmatic, as the economist's time is limited. NICE guidelines are often large and complex, typically covering around 15-20 review questions along a pathway of care, although sometimes many more. Each question may relate to a choice between several different interventions for various subgroups of patients. Modelling will not necessarily enhance GDG decision-making for all questions. For example, if there is clear evidence of a lack of clinical benefit for an intervention, it will sometimes be obvious that it cannot be cost-effective. Alternatively, if there is a lack of evidence of benefit, modelling on the basis of expert

opinion alone might not help the group to reach consensus. Modelling might also be unnecessary if high-quality analyses directly relevant to the decision problem already exist; in a recent UK HTA report for instance. A selective approach to modelling the remaining questions might therefore be sufficient to ensure that the really important economic issues in a guideline are identified and addressed.

However, there are three main risks associated with this selective approach. The first is that important economic issues may be missed or inadequately considered. The existing economic evidence base is usually sparse and patchy, so one cannot rely on published estimates of cost-effectiveness for all of the review questions of interest. For example, a systematic review of economic evaluations of colorectal cancer services found no relevant UK cost-effectiveness estimates for large sections of the care pathway, including surveillance, radiotherapy and end of life care.³⁹ After excluding questions for which economic evidence would clearly not add value, and those covered by sufficient existing evidence, there are usually more cost-effectiveness questions than can be answered with conventional modelling within the resources and timelines of the guideline. Taking the NICE guideline on the diagnosis and management of colorectal cancer (CG131) published in November 2011, the scope specified 15 key clinical questions to be addressed in the guideline.⁴⁰ Of these, the economic plan concluded that one question was already covered by literature, and that cost-effectiveness was not relevant for two questions (one relating to a prognosis and one to support for patients). Of the remaining 12 questions, three were rated as high priority for further economic analysis, four as medium priority and five as low priority. In the event, economic analysis was completed for one high priority topic. This low coverage of economic evidence might have been an inevitable consequence of sparse data and limited modelling resources. However, it is also possible that the expectation that economic analysis will only address a small proportion of guideline questions in a guideline leads to an overly cautious approach, in which difficult (but possibly important) analyses are abandoned.

A second possible risk of the current NICE approach is that GDGs may be forced to make decisions on the basis of inconsistent economic evidence, estimated using different methods, assumptions and data. This could lead to inconsistent application of the cost-effectiveness benchmark to recommendations within a guideline, and also between

guidelines, and between guidelines and other forms of NICE guidance (such as technology appraisals or public health guidance). As an example, in their review of published economic evidence relating to colorectal cancer, Tappenden and colleagues concluded that where economic evidence was available it was 'incongruent and difficult to interpret' between different parts of the same pathway and where multiple analyses existed to address a given decision problem.³⁹ They identified inconsistencies in methodology ('doing things differently') and also in scoping ('doing different things').

The third risk with selective economic modelling is that it may neglect systemic effects and interactions between questions. The sequencing of tests and treatments within the pathway may radically alter costs and health outcomes. In addition, the cost-effectiveness of the options at any node in the pathway may depend on upstream and downstream decisions. For example, the cost-effectiveness of a test depends on downstream treatment decisions, and conversely the cost-effectiveness of a treatment depends on upstream selection of patients. This issue was recognised in the NICE guideline on colorectal cancer, where the GDG chose not to pursue economic analysis for diagnosis, staging or assessment questions because they thought it would be difficult to construct a model structure to take account of downstream events beyond test accuracy.⁴⁰

This combination of sparse and inconsistent published economic evidence and limited capacity for modelling means that guideline recommendations are often not supported by quantitative estimates of cost-effectiveness.

So, there are potential problems with the current NICE approach, but is there a feasible alternative? Alan Williams made a radical suggestion in his 2004 OHE lecture:

"I think that guideline development needs to be strengthened from the outset by injecting into the process a strong dose of decision-analytic expertise, so as to ensure that the whole territory is mapped out in a systematic way, rather than leaving the creation of a comprehensive flowchart until later, when all the bits and pieces on which we have more information have been sorted out..."

"To do this we need not only a large-scale map of what to do at particularly tricky junctions, but also a small-scale map of the entire system covering all the relevant highways and byways, and estimating the traffic flows along each..."

“An impossible task’ did I hear someone mutter? If creating such a map is horrendously complicated, it is because reality is horrendously complicated, but if traffic analysts can do it, surely the health care analysts can do it too! Indeed, the more complex the reality is, the more dangerous it is to rely on intuitive short-cuts rather than careful analysis.”⁸

Examples of full pathway models

A number of ‘generic’ models have been designed to provide a platform for evaluation of a range of interventions for a defined patient group, most notably in the areas of cardiovascular and metabolic disease.⁴¹⁻⁴³ These may be useful for situations where decision makers want a model that can be adapted over time to evaluate emerging technologies or to incorporate new evidence.⁴⁴ One well-known example is the pioneering CHD Policy Model.^{45;46} This was designed to estimate CHD incidence, prevalence, mortality and related resource costs across a population. It used a compartmental state-transition modelling technique, similar to that used for modelling infectious disease dynamics, in which the progress of groups (rather than individuals) is tracked over discrete intervals of time.

Another example is the Department of Health funded Coronary Heart Disease (CHD) model, which used Discrete Event Simulation (DES) to estimate costs and health outcomes of a defined diagnostic and treatment pathway across a population with CHD.^{47;48} Outcomes were determined for simulated individuals through random sampling of the time to CHD events (unstable angina and myocardial infarction) and death (CHD related and other). The simulation included an explicit model of the care pathway, coding the sequence of tests and treatments that individuals would receive, conditional on their characteristics and histories. The pathway was of a similarly broad scope to that in many NICE clinical guidelines, and this example illustrates well how Alan Williams’ vision of a map of an entire guideline might be operationalised.

More recently, the Department of Health funded the development of a guideline-like clinical care pathway model for colorectal cancer.⁴⁹ This also used DES to model current practice, following patients from initial presentation with suspected colorectal cancer through to end-of-life care. The simulation model was then used to provide a baseline for estimation of the cost-effectiveness of a range of potential (largely hypothetical) service developments.

Building on the work of Pilgrim *et al*⁴⁹, Tappenden *et al* later developed a methods framework for developing and using Whole Disease Models to inform resource allocation decisions in cancer.⁵⁰ This methods framework was then applied to inform the development of a Colorectal Cancer Whole Disease Model to examine its potential value in supporting economic analysis within the NICE Clinical Guideline on the Diagnosis and Management of Colorectal Cancer (CG131). Whilst the model was not used directly to inform guideline recommendations, the Whole Disease Model was capable of providing a platform for the economic analysis of eleven of the fifteen guideline topics, compared with only one with the conventional approach. The Colorectal Cancer Whole Disease Model required around 12 months development time however it should be recognised that the authors had considerable previous experience in developing models of colorectal cancer interventions.

Risks and benefits of pathway modelling

The idea of building a model of the full patient pathway to serve as a foundation for economic evaluation in NICE guidelines is attractive.

“In an ideal world, we could develop a single model for a whole disease pathway from diagnosis, incorporating all the different decision points along the way. Looking at a condition in this holistic manner would help to ensure the whole care pathway recommended in the guideline represents the most cost-effective use of resources.”³⁷

A full guideline model has the potential to provide a coherent framework for economic evaluation of a wide range of decision problems within a guideline, ensuring that all analyses are based on a common set of methods, assumptions and data sources. In addition to straightforward comparisons of alternative interventions at an individual node in the pathway, a full guideline model could be used to look at the sequencing of interventions, and also to explore interactions between interventions across different parts of the pathway. Once developed, a full guideline model could be reused to consider other related questions or to incorporate new evidence.

However, this is an ambitious vision. There are technical and practical barriers to the creation of the type of large and complex model that would be needed to cover the wide scope of most NICE guidelines. The general advice in modelling is that simplicity is an

advantage, and that the model structure should be as simple as possible whilst addressing the decision problem and reflecting the nature of the disease and the health care context.⁴⁴ There are certainly potential disadvantages with complex models, as they are likely to be more prone to verification (programming) errors, more difficult to validate, and more difficult to explain to decision-makers than simple models. However, it should also be recognised that “more complex areas require models that respect complexity”.⁵¹ Thus, full guideline models might need to be complex to properly reflect the complexity of guideline pathways. This depends, though, on the extent to which the real-life pathway is interconnected, such that health outcomes and costs in one part of the pathway depend on what happens in another part of the pathway. If the pathway can be segmented, without too serious a loss of realism, it might be safer and more efficient to build several smaller models rather than to attempt to represent them as a whole. Inevitably, the pathways represented in clinical guidelines are always partial reflections of the meta-pathway of the NHS, where patients have multiple diseases and move across the artificial boundaries of guideline demarcations.

Related to the question of complexity, is the choice of modelling technique. It has been argued that whilst individual-level simulation approaches (such as DES) provide greater flexibility than aggregate approaches (such as decision trees or Markov models), they also require specialist skills and may take longer to develop. This view was supported by a study that compared parallel development of a DES and a Markov model to evaluate the cost-effectiveness of alternative adjuvant therapies for early breast cancer.⁵² A contributory factor to this was time spent in understanding how to use DES to model this problem. Additionally, development of a DES model might sometimes be quicker for a larger decision problem, because of the so-called ‘curse of dimensionality’.⁵³ To represent very large decision problems, with multiple subgroups of patients and treatment pathways, aggregate models can require a huge number of health states. For example, Weinstein’s CHD Policy model stratified patients into 5,400 different subgroups, on the basis of differing risk factors.⁴⁵ They then struggled with the problem of how to incorporate coronary angioplasty, which would have doubled this number.⁴⁶ DES has some technical advantages in such cases, since DES can model individual patients and therefore enable them to carry information about their characteristics and history. This can enable a more compact representation of a

heterogeneous mix of patients and complex sequences of decisions and chance events. Thus the simplicity of a model is a function of the size of the decision problem rather than the modelling technique.⁵¹

Similarly, it is sometimes said that data requirements are greater for complex models compared with simple models⁵¹, or for individual-level simulations than for aggregate models.⁵³ However, this is not necessarily true, as data requirements relate more to the size of the modelled problem than to the model structure or technique. For example, the parallel DES and Markov models of adjuvant therapy for early breast cancer mentioned above needed similar data inputs.⁵² Whilst collapsing the number of health states may create a simpler model structure, it is still necessary to estimate weighted means for the transition probabilities, costs and health outcomes for the new combined states.

In addition to these technical issues, there may be wider implications of adopting a more holistic approach to modelling in NICE guidelines. On the positive side, it is possible that the more analytic approach to mapping out the pathway that would be required from the outset could improve evidence collection or guideline decision-making. For example, it might help to define the key clinical questions for review, or it might help the GDG to put this evidence into context. But there are potential dangers. GDG time is limited, and unless they can be seriously engaged with understanding and defining the model structure and data inputs, and in interrogating and interpreting its results, the model will not have credibility and will not influence GDG decisions.¹² Similarly, the ability of external stakeholders to understand and critique the findings might be compromised if modelling methods are too complicated.

The balance between these possible advantages and disadvantages of full guideline modelling is unknown. There is limited evidence to assess whether this approach is feasible, given the practical constraints of resources and timelines for NICE guidelines. It is also uncertain whether the investment will succeed in delivering greater availability or coherence of cost-effectiveness evidence to support guideline recommendations. However, it is certainly plausible that once a full guideline model is developed, it could provide significantly greater insight and ongoing support for decision making across clinical guideline updates and related technology appraisals.

The MAPGuide project

Aims and objectives

The motivation for the Modelling Algorithm Pathways in Guidelines (MAPGuide) project was to test the feasibility and potential usefulness of modelling entire care pathways for NICE clinical guidelines. These models are hereafter referred to as ‘full guideline models’. The aims set for the study were:

- To investigate the feasibility of modelling pathways recommended in NICE clinical guidelines to estimate associated patient flows, health outcomes and costs.
- To illustrate how such models can be used as a basis for assessing the incremental cost-effectiveness of possible variations in the care pathway.
- To use this approach to estimate the value of updating selected topics within the guidelines.
- To compare the update priorities obtained from formal modelling with those elicited during the routine NICE guideline review process.

In order to achieve these aims, we set six key objectives (see Box 5).

Box 5. Project objectives

1. Select two NICE guidelines to serve as illustrative examples.
2. Collate suggestions for topics that could be included in future updates of the guidelines from review documents published on the NICE website.
3. Ask stakeholders for each guideline to rate the suggested topics in terms of priority for inclusion in an update.
4. For each guideline, build a simulation model of the current recommended pathway.
5. Adapt the models to estimate the cost-effectiveness of possible changes to pathways related to the possible update topics.
6. Feedback the results from step 5 to the people consulted in step 3D, and invite them to reassess their ratings of priorities for update.

Background to the project

The project was funded by the Medical Research Council (MRC) and the National Institute for Health Research (NIHR) under their Methodology Research Programme call for research

to underpin NICE decision making. This scheme was intended to fund research into methodological questions of direct relevance to NICE, and followed a scoping study to identify and prioritise topics. One of the highlighted topics was: “assessing the cost effectiveness of 'long' or complex diagnostic/treatment pathways”.⁵⁴ Projects were expected to be completed within two years, to provide rapid feedback to inform policy.

The MAPGuide research team included NICE and NCC staff, as well as academic health economists and simulation modellers. The team has expertise in guideline methodology and systematic reviewing (PA, MW), economic evaluation (MTB, PK, JL, IM, AM, FR, PT, SW, DW), and simulation modelling (JE, PT, ST). Members of the team also have experience of working in the NICE clinical guidelines programme in various capacities: as technical members of guideline development groups (JL, IM, AM, MW, SW, DW); as senior NCC staff supervising the work of technical teams (MW, DW); and as members of the internal NICE guidelines team advising on methodology (PA, PK, JL, FR). The team also has experience of working on NICE technology appraisals (JL, AM, PT).

The project consisted of three main strands of research, which were led by different teams of researchers. Identification of the potential update topics and the survey of stakeholders were led by CC and MW with advice from other members of the project team who were not directly involved with the modelling. Development and application of the simulation models were led by two teams of researchers. The prostate cancer modelling team was based in the London School of Hygiene and Tropical Medicine (SW and AM) in collaboration with the School of Health and Related Research within the University of Sheffield (PT). The atrial fibrillation modelling team was based at Brunel University (JE, MTB, AA, JL). A Project Management Committee, comprising all collaborators and researchers met regularly and oversaw the work.

Rationale for the study design

Two NICE guidelines were selected as case studies to test the feasibility of the full guideline modelling approach. The guidelines were chosen using pre-defined criteria, which included the existence of a relatively well-formulated clinical pathway, which we believed to be a pre-requisite for full guideline modelling. The study therefore represented an attempt at a ‘proof of concept’ that the full guideline model approach could work for selected NICE guidelines, rather than a test of whether it would work for all NICE guidelines.

The idea of using two case studies, rather than one, was to test whether the full pathway modelling approach could work for different types of guidelines, and to explore whether different modelling teams would adopt different modelling approaches. The models were developed by the two teams of analysts who worked separately, but came together to discuss technical and practical issues.

When designing the project, the research team was conscious of the challenging deadlines and resource constraints of 'live' guideline development. The team was also aware that elicitation of an agreed pathway at the beginning of guideline development has proved difficult in the past – topics are usually referred to NICE precisely because there is high uncertainty or disagreement about what is, or should be, standard practice. It was thought to be too risky to test the approach within the real guideline development process. The team therefore decided to first test the feasibility of applying the full pathway modelling approach to two published NICE guidelines. This meant that we started with relatively well-articulated pathways and existing reviews of evidence, which provided a baseline for modelling. If this did not work, attempts to develop such models for new NICE guidelines would be unlikely to succeed. However, to provide a realistic test of the logistics of full pathway modelling, the resources available for developing and using the models were similar in magnitude to those available to NCCs for health economics in a standard NICE guideline: 9 months of analyst time over a period of 18 months for each guideline.

The two case studies were chosen from a list of published guidelines due for three-year review by NICE to determine whether they should be updated. This was intended to provide a convenient opportunity to elicit some questions about possible variations to the pathways in the published guidelines that might potentially be included in a future update of the guidelines. These potential update topics were meant to provide a test for the modellers to assess whether they could adapt their baseline models to address some real cost-effectiveness questions.

We also conducted a survey of stakeholders to elicit their opinions about the relative importance of the potential update topics. This was intended to provide a comparison for the model results, to assess whether they might add value to current methods for selecting update topics. Our original plan was to report a summary of the model methods and results to the stakeholders in a second round survey, and to ask for them if these results would

have changed their prioritisation of the topics (Objective 6, Box 5). However, in the event we were not able to complete this final step of the research plan. This was because development of the models took longer than we had anticipated, and the NCCs started to update the two guidelines that we had chosen as case studies earlier than initially scheduled. This meant that by the time that we had obtained results from the models the development process for updating the two guidelines was already underway. Conducting a second survey at this time could have been disruptive for the NCC and NICE, as stakeholders might have confused our research findings with outputs from the real update. We therefore abandoned the second rounds of the stakeholder surveys. Instead, we simply compared the relative importance attached to the selected potential update topics by stakeholders in our first round survey with the implied importance of these topics based on the results of the model analyses.

The separation of the team identifying potential update topics and conducting the stakeholder surveys from the two modelling teams was intended to avoid bias. The modelling teams did not influence the choice of potential update topics, and were not told what topics had been chosen until after the design of their baseline model had been agreed. This prevented knowledge of the topics influencing the modellers' decisions about the model design, to provide a more robust test of the flexibility of the models to address a range of topics.

Structure of the report

The next chapter provides an overview of the study design and methods. Detailed methods and results for the three main strands of work are reported in the following chapters:

Chapter 3. Stakeholder Surveys

The identification of potential update topics and the survey of stakeholders.

Chapter 4. Prostate Cancer Model

Development of the baseline simulation model and use of the model to investigate possible update topics for our first case study of prostate cancer.

Chapter 5. Atrial Fibrillation Model

Model development and analysis for our second case study of atrial fibrillation.

In the discussion (Chapter 6) the findings across the three strands of research are summarised. We also discuss the strengths and limitations of the study, highlight implications for modelling in NICE guidelines and make research recommendations.

Chapter 2. Overview of methods

This chapter sets out an overview of the methods used in the study.

Selection of case studies

We selected two published NICE guidelines as case studies to test whether the full guideline modelling approach could work. In order to allow sufficient time for modelling within the two-year study period, we considered guidelines due for an update decision by NICE between January and September 2011. This resulted in a list of 17 guidelines that we could have chosen for case studies (see Appendix 1).

The criteria for selection of the case studies defined in our project proposal were:

- Existence of a relatively well-formulated pathway in the current guideline.
- Important topics likely to be updated, so that the models would be likely to have future value in a real update of the guidelines.
- Guidelines for different patient groups or disease areas, likely to present different challenges for the modellers.
- The presence of uncertainty or controversy over which topics should be updated.

The project management committee discussed the options in relation to these criteria (see Chapter 3), and chose the following case studies:

Prostate cancer (CG58) ⁵⁵

This guideline was developed by the National Collaborating Centre for Cancer (NCC-C), and published in February 2008. It was agreed amongst the project team that this guideline has a reasonably clear, well mapped-out pathway with good potential for modelling. After consultation with the NCC-C and a clinical expert, it appeared that an update was likely.

Atrial fibrillation (CG36) ⁵⁶

Developed by the National Collaborating Centre for Chronic Conditions (now the NCGC), this guideline was published in 2006. This guideline also had a clear pathway, with strong potential for modelling. The NCGC reported that there was a fair likelihood that the guideline would be updated.

Identification of potential update topics

The review decisions for the prostate cancer and atrial fibrillation guidelines were published on the NICE website in July 2011 and December 2011, respectively. One researcher (CC) read the review decision and related documents, and collated a list of topics that had been suggested for inclusion in a future update. This list was checked by a second researcher (MW), who is very experienced in systematic reviewing and guideline development. A shortlist of topics for inclusion in the stakeholder surveys and for modelling was agreed by members of the research team who were not involved in developing the models.

Stakeholder survey

Surveys were conducted with registered stakeholders for the two guidelines to elicit their opinions about the importance of the selected potential update topics. Participants were presented with a short summary of the potential topics and then asked to rate each in terms of importance (using a Likert scale), and also to rank them in order of priority for inclusion in a future update of the guideline. Results were summarised in the form of simple descriptive statistics and graphs. The survey methods and results are described in detail in Chapter 3.

Model development

Defining the scope and boundary of the base case models

The research team had to agree some general principles to define the scope and boundaries for the base case models (Box 6). These principles were chosen to ensure that the base case models would provide suitable bases for assessing the cost-effectiveness of possible changes to the guideline recommendations.

Model design

The modelling teams began the process of model design with some background reading to familiarise themselves with their guideline and current issues in the field. They carefully reviewed the published guideline documentation, including the Full Guideline and the Quick Reference Guide (QRG). They also conducted rapid searches to identify other related guidance and key sources of information about their topic. This included NICE guidance and HTA reports, published economic evaluations, guidelines from other national or international bodies, and Cochrane reviews. Ideas for potential model structures and sources of data were identified from these sources.

Box 6. Principles for defining base case models

1. Follow the same scope as for the published guideline. This defines which patient groups, interventions and comparators are to be included or excluded from the model.
2. Reflect as far as possible the pathways recommended in the current guideline, rather than actual practice in the health service, which might vary.
3. Current NICE Technology Appraisal recommendations within the scope of the guideline should be incorporated in full.
4. Pathways for other related NICE guidelines should not be modelled explicitly. For example, in modelling the NICE prostate cancer guideline we decided not to attempt to cover the diagnosis and treatment of benign prostatic hyperplasia, which is addressed in another NICE guideline ¹.
5. Model parameters should be derived from evidence from the original guideline, or from more recent sources identified by rapid reviews or expert advice.
6. Costs and health outcomes should be estimated for an incident cohort of patients over a lifetime time horizon.
7. The starting cohort should reflect a realistic mix of characteristics for patients entering the care pathway.
8. The NICE Reference Case for economic evaluations should be followed ².
9. Uncertainty over model parameters should be incorporated through Probabilistic Sensitivity Analysis. Deterministic sensitivity analysis should be used to explore key structural uncertainties, where appropriate.

Model design broadly followed the phases of ‘problem-oriented’ and ‘design-oriented’ conceptual modelling ^{57;58}: starting with the development of an understanding and description of the relevant health services and disease processes; and followed with the specification of a structure for the applied simulation model and the required information. In practice, there was some iteration between these phases.

Two problem-oriented models were developed in each case study:

- **A service pathway model, which** details the recommended sequence of tests and treatments defined in the guideline. It shows the health services that patients would

receive conditional on their characteristics, if the guideline were to be fully implemented.

- **A disease process model, which details** how patients' health status or risks of events changes over time, conditional on their characteristics and the health services that they receive. This provides the underlying 'engine' that drives patients through the clinical pathway, and is determined by a theory of the natural history of the disease and the way in which treatment effects are expected to interact with that natural history.

The service pathway models were developed following detailed examination of the guideline documents, and were then checked with clinical experts. To support modelling, the flowcharts representing the pathway had to be much more detailed than the 'algorithms' in the QRG version of the guidelines. Some of the ambiguities and discontinuities in the QRG algorithms could be resolved by examination of the precise wording of recommendations, and other text in the Full Guideline document – particularly the 'from evidence to recommendations' sections. The modelling teams resolved remaining uncertainties through discussion with clinical experts.

The disease process models were developed in parallel with the service pathway models. They were designed following review of related published models, descriptions of disease epidemiology (aetiology, progression and prognosis) from the guideline and other background documents, review of outcome measures in the clinical effectiveness data, and discussions with clinical experts. An important factor in finalising the structure of the outcomes models was data availability: including information about baseline risks, treatment effects and quality of life.

Data identification and selection

Parameters required for the models included:

- Disease epidemiology: incidence and prevalence of the condition, risks of adverse events, rates of disease progression, and mortality rates.
- Diagnostic accuracy (e.g. sensitivity and specificity) for any tests in the pathway, including tests used for 'screening', 'identification', 'diagnosis', 'staging', 'assessment', and 'monitoring'.
- Clinical effectiveness of any treatments included in the pathway.

- Quality of life (utility) impact of disease states, events, and treatment side effects.
- Costs of tests, treatments and ongoing care, costs.

In addition, to reflect patient heterogeneity, estimates of relationships between the above parameters and individual patient characteristics were required. These relationships may be in the form of discrete subgroups or continuous covariates. The characteristics can include socio-demographic factors (age and sex), clinical factors (stage or severity of disease) and history (existing co-morbidities or treatments received).

Model parameters were estimated from a variety of sources, obtained from information available in the original guideline, supplemented with new evidence identified from rapid reviews of the literature or from expert opinion. We sought to use the best available sources of evidence, but could not conduct our own systematic reviews, as this was beyond the scope of this project. Where possible, we relied on reviews from the NICE guideline, or from recent high-quality systematic reviews as the source of effectiveness evidence (e.g. Cochrane reviews, HTA reports, or assessment reports for NICE Technology Appraisals). However, it is important to note that the results presented below are not all based on full systematic reviews and that they have not been informed by an expert clinical guideline group. They are intended to be indicative of priorities for full evaluation in a guideline update, and should not be used to inform clinical decisions.

Model implementation

The full guideline models were implemented using Discrete Event Simulation (DES) technique that represented individual patients as entities.⁵⁹ This provided a flexible and relatively compact format for mapping the complicated guideline pathways and predicting outcomes for heterogeneous patient populations.

The models begin with a cohort of patients (the simulation 'entities') with a defined set of personal characteristics ('attributes') at the point of entry to the pathway. The models then follow patients through the care pathway, applying specified rules which dictate the route that patients take as a function of their attributes. These rules may be deterministic (e.g. patients below age 60 receive treatment A, those aged 60 or older receive treatment B) or probabilistic (e.g. 40% of patients receive treatment A, 60% treatment B). For the latter, probability parameters are combined with Monte Carlo (random) sampling to determine the

patient's route through each part of the model. The times to key events (e.g. disease progression, onset of complications or mortality) are sampled for each individual at model entry, and modified as patients progress through the pathway and receive treatments, or if they experience other events. Time to event estimates are based on Monte Carlo sampling from survival functions (Weibull, exponential etc.) fitted to reflect the individual's risk⁶⁰. When sampling time to event values, care is needed to account for 'competing risks': where one and only one of a mutually exclusive set of events can occur.⁶¹ Care is also needed to appropriately modify time to event estimates when things change, for example when someone with atrial fibrillation starts anticoagulation treatment their risk of thromboembolism falls and time to event rises. Individuals' attributes are also updated over time, as they receive different types of healthcare intervention and as they experience key events, as defined by the conceptual models.

The models were programmed using SIMUL8, a dedicated DES package. This was selected as it is generally considered as one of the easier simulation packages to learn, while providing appropriate modelling complexity, excellent experimentation support and has the ability to publish models on the World Wide Web. It has also been used within the NHS, as part of a public private partnership arrangement between the software developer (SIMUL8 Corporation) and the NHS Institute for Innovation and Improvement.⁶²

Verification and validation

The modelling teams checked for errors and inconsistencies throughout model development, following best practice for quality assuring simulation⁵⁹ and decision analytic models.²⁴⁻²⁶ The models were verified internally (to ensure correct programming) and validated (to ensure consistency with expected results – for example, that survival times and levels of service use are realistic). In addition, each of the models was reviewed by an experienced modeller with expertise in DES, who worked with the teams to ensure that any identified errors or inconsistencies were corrected.

Cost-effectiveness analysis

Calculation of base case results

In the base case model, health effects (QALYs) and the costs of interventions and disease-related care were accumulated for simulated individuals (and whole cohorts) as they

progressed through the pathway and disease states, until death. To account for time preference, costs and health outcomes were discounted to the point of model entry, using a continuous discounting approach.⁶³ The results of a defined pathway for each simulated patient i were therefore collected as discounted lifetime sums of costs C_i and effects E_i (QALYs).

In analysing the results of individual-level (micro-simulation) models it is essential to take account of three ways in which model outputs can vary⁶⁴:

- **Patient heterogeneity**, which reflects how model outputs differ across individuals with different characteristics. Within the population of interest (patients entering the pathway), there is a joint probability distribution over some set of initial attributes \mathbf{X} which are functionally related in the model to the outputs $\mathbf{Y} = [C \quad E]$.
- **Parameter uncertainty** ('second order uncertainty') results from uncertainty over values of model input parameters arising from inevitably imperfect knowledge. This uncertainty is represented through a joint probability density function over some set of input parameters \mathbf{B} which are related through the model to the outputs \mathbf{Y} .
- **Stochastic uncertainty** ('first order uncertainty') reflects how outputs for individuals can vary in the model due to chance. This arises because of the stochastic (Monte Carlo) sampling of events and outcomes for individuals. Thus, results may differ for two individuals with identical starting attributes \mathbf{X}_i and a given set of input parameters \mathbf{B}_j .

We conducted a probabilistic analysis^{65,66} of our models to estimate the expected cost \bar{C} and effect \bar{E} across a representative but heterogeneous population of patients treated according to the defined pathways, *and* to estimate the uncertainty around these outputs. This required a nested iteration to integrate over both stochastic and parameter levels of uncertainty: an outer probabilistic sensitivity analysis loop where N sets of input parameters were drawn ($\mathbf{B}_j, j = 1, 2, \dots, N$); and an inner individual-level loop where, for each set of input parameters, n sets of patient characteristics were drawn ($\mathbf{X}_i, i = 1, 2, \dots, n$), and the model was run to calculate the results for each patient ($\mathbf{Y}_{i,j} = f(\mathbf{B}_j, \mathbf{X}_i) + \varepsilon_i$). Results were averaged across the individual-level iterations $\bar{\mathbf{Y}}_j = \sum_{i=1}^n \mathbf{Y}_{i,j}$, and the distribution of the $\bar{\mathbf{Y}}_j$

($j = 1, 2, \dots, N$) used to estimate overall mean results ($\bar{Y} = \sum_{j=1}^N \bar{Y}_j$) and to characterise uncertainty around these results. The choice of the number of probabilistic iterations (N) and the number of individual patients per iteration (n) was made through experimentation: by gradually increasing n until the \bar{Y}_j were stable, and then gradually increasing N until \bar{Y} was stable.

The above process describes how results were derived for one defined pathway – starting with the modelled version of the current guideline recommendations (the ‘base case pathway’). To use the models to conduct cost-effectiveness analysis, the simulation model was then adapted to reflect a range of alternative strategies. Each strategy consisted of one or more changes to the service pathway and/or changes to the model inputs. The alternative versions of the model were run separately, and the results were compared in an incremental cost-effectiveness analysis. To minimise unnecessary variation between the strategies, the individual patient samples and population parameter values that did not differ between the strategies were held constant for each probabilistic iteration j .

Modification of base case model for update topics

As described above, members of the research team not involved in the modelling drew up a shortlist of topics for each model. The shortlisted topics each related to some possible changes to the current pathway, including:

- substitution of different tests or treatments at given points in the pathway;
- changes to patient eligibility criteria or thresholds for tests or treatments;
- different sequencing of tests or treatments and/or
- addition of tests or treatments as an extra step in the pathway.

In addition to the list of topics, sources of new evidence that might support changes to the guideline pathway were identified from the review documents.

After development of the base case version of their models, the modelling teams were given the short list of topics and summary of related new evidence. The modelling teams then attempted to modify their model to represent alternative recommendations that might possibly result from an update of each topic. The modifications ranged from simple changes to input parameters, to minor rewriting of sections of code. We did not attempt any substantial structural changes to the code. Where necessary, we sought additional evidence

to support cost-effectiveness analysis of the topics. As noted above, it was not possible to conduct systematic reviews within the constraints of this project, as this would have required the methodological and subject expertise of the full guideline development process, and our intention was to investigate and illustrate modelling methods, rather than to derive recommendations for clinical practice. The results presented are indicative of the potential value of updating aspects of a guideline, based on the level of reviewing and consultation currently used by NICE and the NCCs when reviewing guidelines for update.

The teams were asked to try to model all of the topics on their shortlist, but as time for the analysis was limited they were invited to prioritise.

Incremental analysis

The models were re-run for each pathway modification, and the same sets of (discounted lifetime) cost and QALY results were accumulated as for the base case model.

Each variation was compared within a full incremental analysis either in terms of Incremental Cost-Effectiveness Ratios (ICER), or using an incremental net benefit (INB) approach. For the ICER analyses, options that were subject to simple or extended dominance were ruled out of the analysis, and ICERs calculated for each remaining option:

$$ICER_{k,k-1} = (\bar{C}_k - \bar{C}_{k-1}) / (\bar{E}_k - \bar{E}_{k-1})$$

where \bar{E}_k and \bar{C}_k are the expected health outcomes (QALYs) and costs under strategies k; and \bar{E}_{k-1} and \bar{C}_{k-1} are the expected health outcomes and costs, respectively, under the next most expensive non-dominated strategy. Results were compared against a cost-effectiveness hold of λ , which was set to the more conservative, lower limit of the range that NICE suggests to its advisory bodies: £20,000 per QALY. The strategy with the highest ICER below the threshold of λ represents the most cost-effective option.

For some analyses an equivalent INB approach was more convenient – particularly where there were a large number of strategies to compare. The INB is defined as:

$$INB_{k,b} = \lambda(\bar{E}_k - \bar{E}_b) - (\bar{C}_k - \bar{C}_b)$$

In this case, each strategy (k) is compared against the base case strategy (b). A positive incremental net benefit result suggests that pathway k is more cost-effective than

the base case pathway b (at the NICE conservative threshold of £20,000 per QALY). The strategy with the largest INB is the most cost-effective of the strategies tested.

The probabilistic results were used to provide an estimate of decision certainty for each comparative result. We calculated the proportion of probabilistic iterations for which the incremental net benefit statistic was positive, $p(INB_{k,b} > 0)$. This is an estimate of the probability that pathway k is more cost-effective than pathway b.

As the analyses were not based on systematic reviews or GDG input, we did not fully characterise the uncertainty surrounding the cost-effectiveness estimates. The ICER, INB and $p(INB > 0)$ results should therefore be seen as preliminary estimates intended to inform a decision about updating the topic, and should not be used to reach definitive conclusions. In addition to this incremental comparison of alternative strategies within each topic, we also sought to compare combinations of strategies between topics to investigate whether there were interactions between them. We had originally intended to also present 'value of information' estimates (e.g. EVPI), as another indication of the potential gains that might be obtained by updating a topic. However, on reflection we decided that these would not provide an appropriate measure of priority for updating. For example, a potential change in recommendation with a high estimated incremental net benefit associated with little uncertainty would have a low EVPI, but would still be an important inclusion in a guideline update. There would also be little to gain from updating a topic associated with high uncertainty (and a high EVPI) unless there was a reasonable expectation that the uncertainty could be resolved by further reviewing and/or GDG discussion.

Usefulness of the full guideline models

Our first method for assessing the usefulness of the full guideline models was to consider the proportion of the shortlisted topics that the modelling teams managed to address within the time available. This is an indication of the appropriateness of the scope and depth of the models, and how easily they can be adapted to answer cost-effectiveness questions.

Secondly, we compared the results of the modelling exercise with the survey respondents' stated priorities over the importance of the shortlisted topics for inclusion in an update.

The modelling teams made judgments about the relative 'economic priority' of the modelled topics on the basis of two key sets of information: i) the estimated probabilities

that the current guideline recommendations are suboptimal, $p(INB_{k,b} > 0)$ for some k; and
ii) the estimated size of potential gain in net benefit from the alternative strategies tested, $MAX(INB_{k,b})$ for all k over a defined population (standardised at 1,000 incident cases in this report). A third method for assessing the potential usefulness of the complex full guideline models was to search for evidence of interactions between the cost-effectiveness of strategies across topics. This would suggest that there are systemic effects that would not be captured by a conventional piecewise analysis of isolated topics.

A final, pragmatic test for the usefulness of the models is whether the collaborating centres and GDGs now working on updates of the two guidelines choose to make use of them. During the course of the project both modelling teams have had discussions with the health economists working on the guideline updates, and agreed to make the models available to them.

Chapter 3. Stakeholder Surveys

Introduction

As part of the MAPGuide study, an online survey was carried out of registered stakeholders of two NICE clinical guidelines (Atrial Fibrillation and Prostate Cancer) to determine their opinions on topics that may potentially be updated within those guidelines.

Aims and objectives

The aim of the survey was to elicit experts' views about the importance of including some suggested topics in an update of the clinical guidelines (Prostate Cancer CG58 and Atrial Fibrillation CG36).

Initially, two surveys were planned: the first of which would be administered before the modelling process in order to elicit opinions on some potential update topics to the guideline, whilst the second survey would be administered after the modelling process to determine whether respondents' views changed in response to feedback about the expected costs and benefits of some potential changes to the guideline pathway based on the models developed in this project. However, in the event only one survey was carried out, as the models were not completed before the NCCs began the guideline update process; to continue the process by sending out a second survey may have caused confusion and so this was abandoned. Instead, estimates of the cost-effectiveness of potential changes to the pathway associated with update topics were obtained from the models, and compared with the stated priorities of stakeholders about the relative importance of updating topics.

As the aim of the surveys was to test the usefulness of the modelling, they were conducted by a researcher (CC) from Brunel University who was not involved in the model development process. Advice and guidance regarding identification of topics and survey development was received from a member of the National Clinical Guidelines Centre (NCC), Royal College of Physicians (MW).

Methods

Ethical Approval

Ethical approval was applied for and received from the university research ethics committees based at Brunel University and the London School of Hygiene and Tropical Medicine. Approval from an NHS ethics committee was not necessary as the participants were not identified on the basis of their status as NHS patients or staff, and no research was conducted on NHS premises.

Identification of potential update topics

Potential update topics for the surveys were identified by one researcher (CC) after review of the following documents, obtained from the NICE website:

- NICE's review proposal and any related consultation documents
- The table of stakeholder consultation comments and responses
- NICE's final review decision and any supporting documentation.

From these documents lists of possible update topics and new evidence relating to those topics were compiled for each guideline. These topics were defined at the level of 'Key Clinical Issues':

"Key clinical issues relate to the effectiveness and cost effectiveness of interventions or tests that are being considered for a given population. These issues should be developed out of a care pathway or a similar analytical framework. They are not the same as review questions, which specify in some detail the particular interventions to be compared and the health outcomes of interest Nevertheless, key clinical issues should be as specific as possible, indicating the relevant population and the alternative strategies that are being considered." (Chapter 2 p22, NICE Guidelines Manual 2009) ³¹

A second researcher (MW) reviewed the list and both researchers then used the following criteria to derive a shortlist of topics:

- All update topics specified in NICE's review decision
- Topics which had substantive support from stakeholders and other experts

To avoid overburdening respondents, it was also decided by the researchers undertaking the survey that the maximum number of topics for each survey would be ten.

Regarding the choice of which topics to include in the shortlist, if both researchers were in agreement, this topic was included. If there was uncertainty over a topic's inclusion, both researchers reviewed the evidence for that topic to determine if there was sufficient information about the update topic to warrant its inclusion, if not it was excluded. To limit the number of topics, some were also excluded if it was deemed the other topics may have more impact on the guideline pathway or had more stakeholder support.

A number of topics were excluded from the prostate cancer survey shortlist including a topic regarding the addition of guidance for the use of an F18 Choline PET CT scan for diagnosis of recurrent disease after radical prostatectomy or radiation therapy. This was recommended in the guideline as an option for patients with biochemical recurrence after negative MRI and bone scans. This topic was excluded as there were limited papers available to confirm its importance as an update topic. Another topic that was excluded was that pertaining to the use of Docetaxel as a first-line treatment option for men with hormone refractory prostate cancer. This particular drug was already included as a recommendation within the current guideline, and a review of the technology assessment of the drug has been postponed until 2013. It was decided to exclude this topic as it is already recommended within its licensed indication within the guideline and these criteria could not be amended.

A number of topics were also excluded from the shortlist for the atrial fibrillation survey including the introduction of a nationwide opportunistic screening programme by integrating manual pulse checks as part of national screening flu programmes or chronic disease management. This was excluded as it was deemed to be outside the current scope of the guideline.

The final topics included in the surveys were agreed upon by the MAPGuide project management committee members who were not involved in the modelling process. Potential update topics chosen for the prostate cancer survey (nine topics) and for the atrial fibrillation survey (eight topics) are listed in Box 7 and Box 8.

Box 7. Potential update topics for prostate cancer guideline

- A. Pelvic radiotherapy with adjuvant hormonal therapy for men with localised prostate cancer
- B. Effective techniques for performing radical prostatectomy.
- C. HDR Brachytherapy in addition to external beam radiotherapy for men with localised or locally advanced prostate cancer.
- D. LDR Brachytherapy in addition to external beam radiotherapy for men with localised or locally advanced prostate cancer.
- E. Degarelix (a LHRH antagonist), for men with advanced hormone dependent prostate cancer (locally advanced or metastatic).
- F. Intermittent hormone therapy versus continuous hormone therapy for men with metastatic prostate cancer.
- G. Radium 223 chloride versus strontium-89 for men with hormone refractory prostate cancer and painful bone metastases.
- H. Intensity Modulated Radiation Therapy (IMRT) and Image Guided Radiation Therapy (IGRT) as an alternative to conventional therapy for men undergoing radiation treatment
- I. Active surveillance in previously unscreened 'low risk' men

Box 8. Potential update topics for atrial fibrillation guideline

- A. Prophylaxis for the prevention of post-operative AF
- B. Anti-arrhythmic drugs as pharmacological cardioversion for people with Atrial Fibrillation
- C. Rhythm versus rate control strategies for persistent AF; updating eligibility of subgroups including those with hypertension, previous MI and congestive heart failure
- D. Treatment for maintaining Sinus Rhythm in people with AF after cardioversion.
- E. Alternative risk factor based scoring systems to estimate stroke and embolism risk
- F. Stratification tools to assess bleeding risk before prescription of antithrombotic medication
- G. Apixaban, rivaroxaban or Dabigatran etexilate (anticoagulants) versus warfarin as thromboprophylaxis for patients deemed at moderate or high risk of stroke or systemic embolism
- H. Catheter ablation for paroxysmal and persistent AF patient

Survey development

The questionnaire was compiled using an online commercial survey tool, Survey Monkey™ (<http://www.surveymonkey.com>). The survey invited participants to rate the relative importance of including each topic in a potential future update of the guideline, using a Likert scale with five options ranging from 'Not Important' to 'Very Important' with an option to choose 'No Opinion'. Participants were also asked to rank the suggested topics in order of preference for inclusion in an update. Free text comment boxes were added to seek qualitative information and reasons for responses. Online links to information sheets were provided. These were developed by the survey researcher (CC) and provided details of the original guideline recommendation and the section of the clinical pathway the update topic would have an impact upon, and specific details of the proposed update topic (new evidence on the topic and comments from stakeholders) which was taken from the guideline review documents.

Pilot Study

A pilot study was conducted to test the adequacy and suitability of the survey. The sample for the pilot was drawn from outside the survey sample. The pilot sample consisted of contacts of the Project Management Committee from other related guideline development groups (GDG's) and clinicians with an interest in the relevant guideline. The pilot sample for each study consisted of three clinicians who were contacts of the Project Management Committee.

An internal pilot was also carried within the Health Economics Research Group based at Brunel University to ensure that the survey would be suitable for lay-persons without clinical knowledge of either prostate cancer or atrial fibrillation.

A small number of changes were made to the surveys after the pilot, including amendments of any grammatical or spelling errors, some changes were made to the layout (for example enlargement of the font size) but no changes were made regarding the chosen topics.

Survey of experts and elicitation of views about potential update topics

The sample for the survey was drawn from the list of registered stakeholders for the guidelines, ex-GDG members and NCC staff who contributed to the development of the guideline and are named in the published guideline on NICE website.

The following numbers of registered stakeholders, GDG members and NCC staff were invited to participate in the surveys:

- Prostate cancer guideline survey, 225 registered stakeholders, 14 GDG members,(total N=239)
- Atrial fibrillation guideline survey, 168 registered stakeholders, 14 GDG members (total N=182)

The list of stakeholders included a wide range of individuals affiliated with organisations with an interest in both guideline topics, including: patient organisations, specialist societies (doctors, nurses and other professions allied to medicine), industry, and other health service organisations.

As the registered stakeholder list is held by NICE, it was not appropriate for the researcher to send an invite to participate in the survey directly to the stakeholders. Therefore, an email (prepared by the researchers and approved by the ethical committee) inviting registered stakeholders to participate in the research project was sent by the NICE Centre for Clinical Practice project manager, along with a Frequently Asked Question (FAQ) sheet providing further information about the study and a link to the online questionnaire. The researcher contacted ex-GDG members, NCC staff and other experts named in the guideline directly with an invite to participate in the study.

No reminders were sent to individuals who did not respond to the initial invitation to participate, as we did not have direct access to the email addresses of stakeholders.

Analysis of responses

Results from the survey were entered into SPSS, and descriptive statistics including simple measures of central tendency (median) and frequency (distribution) results were calculated. Qualitative data obtained from the additional comment fields in the surveys were compiled and analysed for similar themes.

Results

Response Rate

32 persons responded to the atrial fibrillation survey and 27 persons responded to the prostate cancer survey giving a response rate of 19% for atrial fibrillation and 14% for

prostate cancer. The response rate was expected to be similar to the number of stakeholders that responded to NICE's call for comments from registered stakeholders on the review consultation document for both guidelines which listed potential topics that may either be updated or added to both guidelines (21 stakeholder organisations provided comments on the review consultation document for the atrial fibrillation guideline and 27 stakeholder organisations provided comments on the review consultation document for the prostate cancer guideline, a response rate of 13% and 14% respectively).

Missing Data

18 of the 27 respondents to the prostate cancer survey invite completed the survey and all 18 completed both questions. Seven respondents added a comment in the 'additional comments and information' fields.

Not all persons who responded to the atrial fibrillation survey invite completed the survey. 25 of the 33 respondents completed question one (rating question) and 23 of these 25 respondents also completed question two (ranking question). 13 respondents added a comment in the 'additional comments and information' fields.

Two respondents answered question one but not question two; one of these respondents' answers to the rating question differed from the remainder of the sample. They were the only respondent to enter 'no opinion' for all eight topics. The other respondent who did not complete question two rated three of the topics (3, 4, and 5) as being somewhat important and rated the rest as important for question one; this was a similar response to those who completed both questions in the AF survey.

Organisational Affiliation of Respondents

Of the 25 respondents who completed the atrial fibrillation survey, 44% were affiliated to Health Services Sector, 20% to Industry, 12% to charities, 8% to Specialist Societies and 16% to other such as educational and informational organisations. Of the 18 respondents who completed the prostate cancer survey, 50% were affiliated to the Health Services Sector, 17% to Charities and Patient-Led Societies, 17% to Industry and 6% to other such as educational and informational organisations.

Median responses

The median results for question one (rating question) and question two (ranking question) are presented in Table 1 and Table 2 for the prostate cancer and atrial fibrillation surveys respectively.

Table 1. Responses to prostate cancer survey

Topic Description	Question 1 Median rating (n=18)	Question 2 Median rank (n=18)
A Pelvic radiotherapy with adjuvant hormonal therapy	4	5
B Effective techniques for performing radical prostatectomy	5	2
C HDR Brachytherapy in addition to External Beam Radiotherapy	4	4.5
D LDR Brachytherapy in addition to External Beam Radiotherapy	4	4.5
E Degarelix for men with advanced hormone dependent prostate cancer	3.5	8
F Intermittent hormone therapy versus continuous hormone therapy	4	3
G Radium 223 chloride versus strontium-89	3.5	8
H IMRT and IGRT as an alternative to conventional radiotherapy	4	5
I Active surveillance in previously unscreened 'low risk' men	5	2.5

Table 2. Responses to atrial fibrillation Survey

Topic Description	Question 1 Median rating (n=25)	Question 2 Median rank (n=23)
A Prophylaxis for the prevention of post-operative AF	3	7
B Anti-arrhythmic drugs as pharmacological cardioversion for people with AF	4	4
C Rhythm versus rate control strategies for persistent AF	4	5
D Treatment for maintaining Sinus Rhythm in people with AF after cardioversion.	4	5
E Alternative risk factor based scoring systems to estimate stroke and embolism risk	4	4
F Stratification tools to assess bleeding risk before prescription of antithrombotic medication	5	3
G Apixaban, rivaroxaban or Dabigatran etexilate (anticoagulants) versus warfarin	5	2
H Catheter ablation for paroxysmal and persistent AF patients	4	6

Question 1: Rate the importance of including the topic in an update of the guideline: 5 'very important', 4 'important' 3 'no opinion', 2 'somewhat important' and 1 'not important'.

Question 2: Rank in order of importance for inclusion in an update of the guideline: 1 most important to 9 least important.

Distributions of Results

The frequency distributions of rating and ranking responses for each topic are shown in Appendix 2.

Comparison of Rating and Ranking Results

There were some slight differences in the overall results to the two questions (rating and ranking) regarding which topics were perceived as the most important to update. Table 3 and Table 4 display the three topics deemed the most important to update by respondents for both questions.

Table 3 . Top three prostate cancer topics

	Question 1 (Rating)	Question 2 (Ranking)
1	Active surveillance in previously unscreened 'low risk' men	Active surveillance in previously unscreened 'low risk' men
2	Effective techniques for performing radical prostatectomy	Effective techniques for performing radical prostatectomy
3	IMRT and IGRT as an alternative to conventional radiotherapy for men undergoing radiation therapy	Radium 223 chloride versus strontium-89 for men with hormone refractory prostate cancer and painful bone metastases

Table 4. Top three atrial fibrillation topics

	Question 1 (Rating)	Question 2 (Ranking)
1	Apixaban, rivaroxaban or Dabigatran etexilate (anticoagulants) versus warfarin as thromboprophylaxis	Apixaban, rivaroxaban or Dabigatran etexilate (anticoagulants) versus warfarin as thromboprophylaxis
2	Stratification tools to assess bleeding risk before prescription of antithrombotic medication	Alternative risk factor based scoring systems to estimate stroke and embolism risk
3	Alternative risk factor based scoring systems to estimate stroke and embolism risk	Catheter ablation for paroxysmal and persistent AF patients

The results also show that there is a similarity in the individual respondent's answers to question one and two. The prostate cancer survey had 18 respondents, and the results were analysed to determine if each respondent's three top rated topics for question one (rated on a Likert scale from not important to very important) matched with their top three

ranked topics for question two. Eight respondents gave the same response to both questions. For three respondents it was not possible to tell if the responses matched, as their answer to the first question rated all the topics the same (either rated all as most important or as no opinion). One respondent had ranked the topic they had rated the most important as second most important to update, so there was a slight discrepancy between results, and they had rated all the other topics as important so it was not possible to determine any correlation. Six respondents rated and ranked the first two topics they deemed as the most important to update the same, however they gave a different response for the third topic, whereby a topic that may have been either as most important or important to update may have been ranked the least important to update.

Of the 27 atrial fibrillation respondents who completed the survey, only 25 completed both questions. Therefore for two respondents it was not possible to compare their rating and ranking scores. Of the 25 respondents who completed both questions, 12 has similar responses for question one and question two (topics rated most important were also ranked in the top 3). Seven respondents had the same response for both ranking and rating for the first relevant topic but not for the remainder. It was not possible to compare the score for two respondents as for question one they had rated all the topics the same (either rated all as most important or as no opinion).

From analysing the data, it seems there may also have been confusion over the ranking of the topics- where two respondents may have perceived the ranking score (scale 1-8) as going from least important (1) to most important (8) to update. This can be seen with the correlation of data whereby one respondent rated Topic 8 (*Cather ablation for paroxysmal and persistent AF patients*) as the 'most important' to update yet ranked it last, and ranked the topic they deemed the least important as the first to be updated. Also, similarly the other respondent rated Topic 5 (*Apixaban, rivaroxaban or Dabigatran etexilate (anticoagulants) versus warfarin*) as 'most important' to update yet ranked it as the least important (ranked number eight) topic to update.

The difference between the overall top rated and ranked topics and the difference between individual's responses to question one and two may be due to a number of similar factors; one being the number of topics in the survey which may have meant that respondents may not have remembered what rating they had given the topics when they went to rank the

topics (there was a facility to toggle back and forward between the survey questions but this may not have been used). Another reason may be that respondents hold strong opinions on one or two topics but not on all the topics so this could explain dissimilar rankings after the respondent had chosen what they deemed to be the most important topic to update. With the atrial fibrillation survey, there also seems to have been confusion over the direction of ranking scores caused by the fact that the Likert Scale asked respondents to rate topics on a scale ranging from 'Not Important' to 'Important' whereas the ranking scale asked respondents to rank in order of preference on a scale of one to eight (where one is most important) rather than ranking in order of least preference similar to the Likert Scale (where one is least important); this could mean that future surveys that employ two different methods to elicit similar information should ensure that both scales are analogous.

Qualitative Data- Comments from Free Text boxes

The comments received from the respondents for the prostate cancer survey were analysed to determine if they could be grouped into similar themes. However, each comment related to a different topic so it was not possible to group the topics. Box 1 lists all the comments received from respondents to the prostate cancer survey.

Comments received from respondents to the atrial fibrillation survey could be grouped under a number of similar themes; two respondents called for the guidelines to be simple, two others referred to the use of the stroke risk scheme and also called for clearer guidance on catheter ablation. Others were interested in technology assessment for new drugs, while two others also mentioned the importance of screening for atrial fibrillation. Box 2 displays the comments from the survey.

Discussion

The results indicate that most respondents agreed on the important topics to update within the guidelines, although there differences for other topics. A major limitation to achieving the initial objective of the survey was that we were unable to assess if respondents opinions would change after viewed the modelling results. To evaluate the modelling, however the results of the first survey could still be used to compare the topics deemed important by the survey respondents and the topics that the model analysis indicates are likely to be important.

Box 9. Prostate cancer survey comments

“There is no provision for updates; consideration of new advances, techniques; clinical data supporting the use of HIFU as an alternative therapy for men who fall within known selection criteria”

“Greater patient benefit from enhanced recovery programmes for prostate cancer, so still a mystery as to why excluded!”

“Impact of long term hormonal therapy on bone health is an important item that is not being considered”

“There are a number of new drugs about to be licensed for the treatment of advanced or metastatic prostate cancer. The current treatment pathways are unclear and will become more complicated with the introduction of Abiraterone, cablitaxel and the desire to use these sequentially. Also there is patchy adherence to current guidance as some patients are having re-exposure to Docetaxel. It would be helpful to give some idea of the rationale for choosing chemotherapy and hormonal therapy and any factors to be taken into consideration for sequential treatments.”

“The question concerning ‘effective techniques’ for performing radical prostatectomy MUST involve an honest health economics review. In the absence of any clinically significant benefit for vastly more expensive, commercially driven treatment modification argued for by eminent self interest groups, i.e. robotic radical prostatectomy, these treatments must be ruled against unfavourably. The most pressing question, I believe, concerns the safe advocacy of active surveillance in low risk men.”

Box 10. Atrial fibrillation survey comments

“Please keep it simple”

“Last document had complicated pathways-make simpler this time.”

“All these questions relate to specific interventions and none relate to the functional capability of living with AF-the natural inference is the target of treatment is to control treatment rather than promote function”

“The stroke risk scheme needs to be simplified and ideally should simply align with chads. If it doesn't then chads will be used anyway. Removal of ageist statements restricting therapy on the basis of age should be removed unless substantial data can justify them. Catheter ablation requires very clear guidance because a large post code variation exists in the UK because of differing behaviour across PCT's.”

*“AF should be reviewed and updated using CHADS2VASC2 scoring and HAS BLED
www.afstrokerisk.org”*

“New anticoagulants are important but are subject to individual technology appraisals so cannot be assessed vs warfarin in the clinical guideline”

“Atrial fibrillation needs to be a high priority on the health care agenda”

“Unable to comment on pharmacological guidelines. They are less relevant to medical devices company's areas of work”

“It is very important to update the guidelines with up to date available information on clinical trials and approved new alternative drugs used in the management of AF”

“An MTA of all agents available for stroke prevention will be most helpful”

“The cost implications (Budget Impact/Cost-Effectiveness (Cost per LYQ/QALY/DALY) of strategies in AF management alone are critically important and the screening of individuals for AF health checks are also critical. The communication and engagement of patients in their own decisions is also critical. The appropriate consenting of patients for ablations with appropriate and realistic data on ablations is critical also. Ablation long term does not get properly discussed with patients and a failure rate of 60% post first procedure is never discussed.”

“AF screening should be a high priority also”

Chapter 4. Case study 1: Full guideline model for prostate cancer

This chapter presents a case study application of the development of a full guideline model to evaluate multiple decision problems across the prostate cancer pathway.

Introduction

Introduction to the context of the case study

Prostate cancer is the most common cancer in men in the UK.⁶⁷ Every year over 40,000 new cases are diagnosed and just over 10,500 men die of prostate cancer. It is largely a disease that affects older men and is rare below the age of 50. More than 75 per cent of cases occur in men over 65 years of age, with the largest number in men aged between 70 and 75 years old.⁶⁸ The symptoms of prostate cancer can be easy to misinterpret as they are not specific to the disease. They include urgency, difficulty and pain on passing urine. Men with early stages of the disease are likely to have no symptoms at all.

There is no routine screening of men in the UK for prostate cancer⁶⁹, however men are encouraged to seek a consultation with their GP for testing if they are concerned about or are at higher risk of developing the disease. Risk factors for prostate cancer include age, family history (the risk of developing prostate cancer doubles or triples for men with a family history of prostate cancer in a first-degree relative), ethnicity (the incidence of prostate cancer in the UK is highest in black Caribbean and black African men and lowest in Asian men) and diet (diets high in calcium may increase the risk of developing prostate cancer).⁶⁸

Prostate cancer is not always life-threatening. Over the past 10 to 15 years there have been a number of significant advances in prostate cancer management but also a number of major controversies, particularly about the clinical management of men with early, non-metastatic disease.⁵⁵ Radical treatment can result in nerve damage and cause urinary dysfunction, sexual dysfunction and bowel problems which have a significant and lasting impact on quality of life.

Variation in practice across the UK, the significant uncertainties faced by men in making treatment decisions and the considerable impact of prostate cancer on quality of life as well as mortality led to the commissioning of the first clinical guideline on prostate cancer by

NICE in 2005. The guideline covered the key aspects of prostate cancer management from the point of referral into secondary care: diagnosis and staging, observation, radical treatment, salvage treatment, follow-up, hormone treatment and best supportive care.⁵⁵

Aims of the case study

The aim of this case study was to develop a health economic model to cover the scope of the prostate cancer guideline in sufficient depth that it could be used to evaluate various options for service change.

The modelling approach was broadly based on the methodological framework for developing Whole Disease Models set out by Tappenden *et al*,⁵⁰ albeit using a more restrictive model scope which includes only a partial representation of disease natural history.

This case study includes economic analysis of a number of potential topics to update within the guideline, selected using methods discussed in chapter 3. Our aim was to investigate the ability of the full guideline model to address such questions. The results of these analyses are not intended to provide suggestions for new guideline recommendations, as they are not based on up-to-date systematic reviews of clinical effectiveness and they have not been informed by an expert clinical guideline group. Instead the aim of the economic analyses is to indicate topic areas where further investigation is likely to be of value.

Methods

The typical starting point for creating a health economic model involves developing an understanding of the decision problem under consideration and a clear definition of a research question which sets out the full set of relevant decision alternatives to be compared and the basis for the comparison (e.g. incremental cost per QALY gained). Given that one of the key objectives of the case study was to assess the flexibility of having a full guideline model for prostate cancer, the questions that the model would need to be able to evaluate were not known at the outset.

We developed a detailed individual-level discrete event simulation (DES) to evaluate the expected cost-effectiveness of options for the diagnosis, treatment and follow-up of prostate cancer. The model was developed using SIMUL8[®] software (SIMUL8 Corporation, Boston). In line with the current NICE Reference Case,² the model considers health

outcomes and costs from the perspective of the NHS and Personal Social Services (PSS) and simulates key clinical and subclinical events, and the costs and consequences of these, over the remaining lifetime of patients. Costs were valued at 2010-11 prices. All costs and health outcomes were discounted at a rate of 3.5%. The headline model results are presented in terms of the incremental cost per quality adjusted life year (QALY) gained within each guideline topic.

The model development process had four main stages. Firstly, we developed a detailed understanding of the clinical area and represented this using conceptual service pathways models. These conceptual models are intended to be recognisable to men with prostate cancer in the UK and to clinicians working within the NHS. This aspect of conceptual model development was based on a preliminary review of the literature and the existing NICE guideline.⁵⁵ We also developed an understanding of the key clinical events, and later represented these within a model of the disease process. The second stage involved converting our understanding into a model constructed to retain the key events in the clinical pathway, whilst taking into account the availability of evidence and the need for simplifications and assumptions. This took the form of a design-oriented conceptual model which set out the main interactions between the disease and treatment pathways. This latter conceptual model was developed iteratively and was only formalised at a late stage during model development. The third stage involved programming the simulation model. Whilst it has been argued that conceptual model development and implementation should remain largely discrete⁵⁷, the processes of designing and implementing the model overlapped considerably. The final stage involved using the model to assess the potential cost-effectiveness of a variety of options for service change across the prostate cancer pathway.

Preliminary literature review

We conducted a literature review of published economic models of prostate cancer from NICE (technology appraisals) and other HTA bodies and guideline developers. Searches were undertaken across a number of electronic databases (CRD NHS Economic Evaluation Database, CRD HTA Database, NHS Evidence, Cochrane Library and Guidelines International Network database) using general disease and patient group search terms. This search was undertaken as a rapid means of identifying potentially appropriate structures for certain

elements of the model and to identify potentially relevant sources of evidence to inform the model parameters. We did not conduct a formal critical appraisal of the identified economic evaluations nor did we summarise their findings, as we were not specifically interested in the credibility of the results of existing models. Documentation for the current NICE guideline was reviewed (comprising the full NICE guideline, accompanying evidence review, the quick reference guide and the implementation tools⁷⁰⁻⁷³) in detail to ensure that we had a coherent understanding of existing recommendations and the rationale underpinning these, the recommended care pathway and the clinical and economic evidence available at the time the recommendations were made.

The conceptual model

Boundary and scope of the model

The scope of the NICE prostate cancer guideline⁷⁴ was used to define the boundary of the health economic model. Entry and exit rules were defined, based on all current recommendations from NICE including recommendations for men with hormone-refractory disease from the NICE Technology Appraisal TA101 (NICE 2006) . Patients enter the model after having been referred to secondary care by their General Practitioner (GP), either due to the presence of symptoms or due to an elevated prostate specific antigen (PSA) test. Patients exit the model when they die or when they have an event which would fall under the remit of another guideline. For example, although the NICE Prostate Cancer Guideline⁵⁵ refers to the referral of patients with suspected prostate cancer from primary care, this is covered in another guideline⁷⁵ and was thus deemed to be beyond the boundary of this evaluation. A proportion of men who present with elevated PSA will not have prostate cancer but may still undergo further tests and monitoring for prostate cancer, so these patients were necessarily retained within the model boundary.

Conceptual service pathways

A conceptual representation of the clinical service pathways for prostate cancer services in England and Wales was constructed based on the recommendations contained within the NICE 2008 prostate cancer guideline.⁵⁵ This was intended to represent clinical practice *if* the recommendations within CG58 had been fully implemented. It is important to note that the

pathway does not necessarily reflect actual practice in the NHS, as the extent of implementation and compliance with guideline recommendations is likely to be variable.

The NICE prostate cancer guideline has a relatively clear structure in terms of the key disease management areas; diagnosis and staging of disease, monitoring and management options, potentially curative treatment and palliative treatment. However, like most clinical guidelines, it was not designed to cover every aspect of clinical care hence a number of assumptions were required to link individual recommendations into a single “joined-up” pathway. We sought advice from a consultant clinical oncologist who was a member of the 2008 Prostate Cancer NICE GDG and an additional urological registrar to ensure the accuracy and representativeness of the conceptual service pathway.

Figure 1 summarises the conceptual service pathway model; a more detailed version of this conceptual model is presented in Appendix 3. Briefly, patients enter the pathway on referral into secondary care. Patients may have been referred by their GP or by another secondary care physician. Repeat tests are undertaken during the initial consultation and a decision is made whether the patient should undergo a transrectal ultrasound (TRUS) biopsy. If men opt out, or if a TRUS biopsy is not considered necessary, they have regular PSA tests carried out by their GP. Note that although men on GP monitoring will not have a diagnosis of prostate cancer at this point, some may have the disease. For men who do undergo TRUS biopsy, the result (which generates a Gleason score - a marker of cell differentiation or ‘aggressiveness’ of the cancer) is used together with PSA score and clinical disease stage to define a patient’s risk (see Table 5).

Table 5. D’Amico risk classification⁷⁰

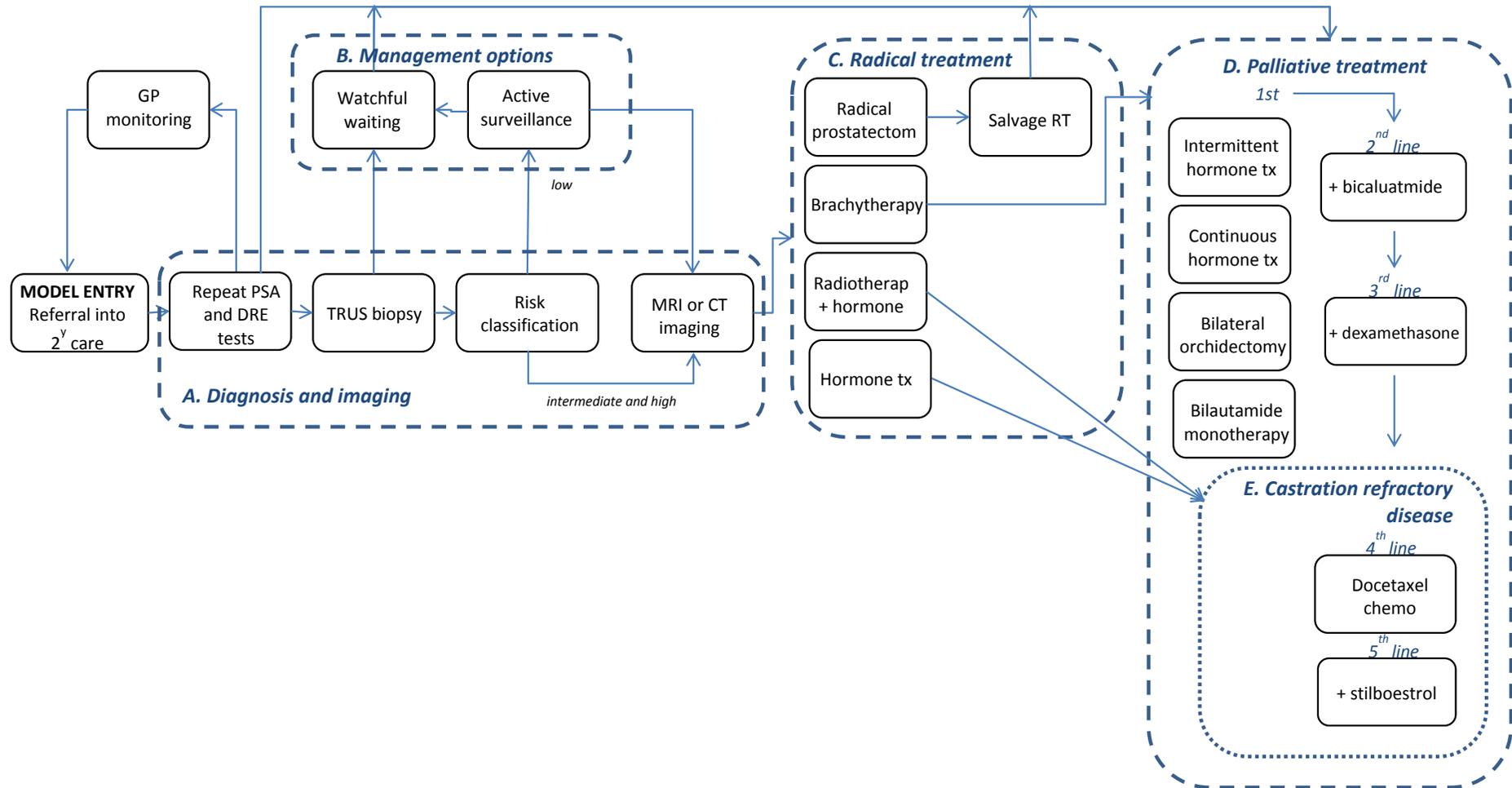
Risk	PSA (blood test)	Gleason (biopsy)	Clinical stage (digital rectal examination)	
Low	<10 ng/ml	≤ 6	T1-T2a	Localised disease
Medium	10-20 ng/ml	7	T2b or T2c	
High	>20 ng/ml	8-10	T3-T4	Locally advanced disease

The “preferred treatment option” for men with low-risk disease who are suitable candidates for radical treatment is active surveillance (AS).⁵⁵ We interpreted this preference as a strict recommendation, ignoring other treatments recommended as possible alternatives.

All men with intermediate- or high-risk disease who are suitable for radical treatment are assumed to receive imaging (magnetic resonance imaging [MRI] or computed tomography [CT] scan) to stage the disease and plan treatment. Radical treatment options include prostatectomy, brachytherapy (for patients with high risk disease only), radical radiotherapy with adjuvant hormone treatment or hormone treatment (in which case men follow the same pathway as for, what we term, 'palliative treatment').

Men who are considered unsuitable for radical treatment or who have a life expectancy of 10 years or less are assumed to receive watchful waiting. This involves regular PSA tests and contact with a urologist in a secondary care setting. If symptoms of advanced prostate cancer develop over this time individuals are assumed to receive palliative treatment. First-line palliative treatment was taken to mean either medical or surgical castration (intermittent or continuous hormone treatment or bilateral orchidectomy) or bicalutamide monotherapy (which may be chosen to retain sexual function at the expense of overall survival). When first-line treatment fails, bicalutamide is added to the treatment regimen (unless the patient has received bicalutamide previously, in which case continuous hormone treatment is offered) and the addition of dexamethasone is given as third-line palliative treatment. When dexamethasone fails, the patient is considered castration refractory.

Figure 1 Summary of the service pathway based on guideline recommendations

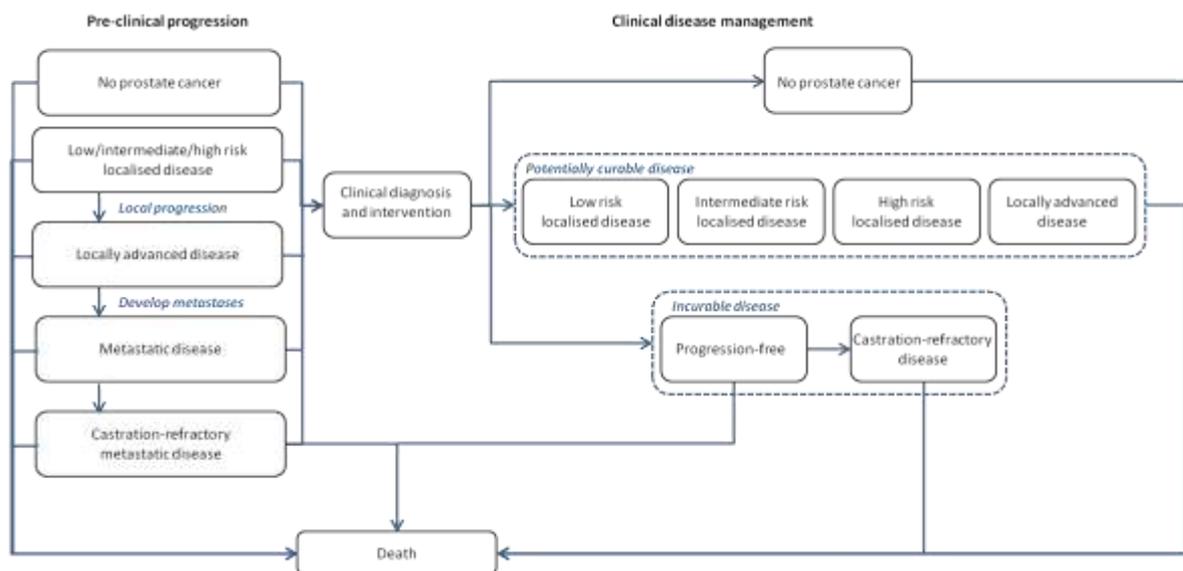


If the patient is considered well enough, chemotherapy is offered as fourth-line palliative treatment, using either docetaxel or mitoxantrone in combination with prednisone or prednisolone.⁷⁶ When chemotherapy fails, patients receive corticosteroids, such as stilboestrol for pain relief. No further active treatments are offered after this time, patients will receive best supportive care.

The disease process model

In addition to the service pathways model, we also developed a conceptual model of the disease process to characterise the key clinical events, risks and subsequent prognosis (see Figure 2). We assumed that prior to diagnosis the underlying progression of prostate cancer follows a consecutive sequence of disease events, depicted on the left hand side of Figure 2. Men without prostate cancer are only at risk of death from other causes. Men with localised prostate cancer are assumed to only develop metastases if they first have local progression. The NICE prostate cancer guideline recommends that a clinically meaningful relapse should be established before starting palliative treatment. Owing to the absence of reported evidence on documented relapse, we assumed that biochemical relapse after radical treatment is a proxy for local progression. Similarly, we assumed a patient cannot die of prostate cancer without first developing metastases.

Figure 2. Disease process model



The central distinction in the clinical management of the disease (depicted on the right hand side of Figure 2) is between patients with disease that is potentially curable, and those with disease that is not. The distinction between localised disease and locally advanced disease is assumed to be less significant since treatment options for patients with locally advanced disease mirror those offered to patients with high risk localised disease. The aim of treating patients with incurable disease is to slow the progression of the disease and to prevent it becoming castration-refractory. Patients with castration-refractory disease may be treated with chemotherapy, which is also intended to slow the progression of the disease. Again due to limitations in the available evidence on documented relapse, we assume equivalence between biochemical relapse and local progression.

Final model design

The final model structure did not fully mimic the conceptual service pathways model described above. The main reason for this was that the 2008 NICE guideline relies heavily on PSA score as an indicator of underlying disease progression and as a trigger for events such as follow up tests and changes in treatment. Some evidence was available on initial PSA and PSA changes over time according to initial diagnosis (which we used in the GP monitoring section of the model), however we did not find evidence to link these changes in PSA to changes in treatment or risk of progression over time. As a consequence, we were unable to use PSA to fully drive changes in patients' underlying disease and the treatment pathways that patients would follow. Instead, we assumed that the natural history of the disease follows a linear series of conditional transitions from local progression to metastases to death from prostate cancer (see figure 2). We also assumed that patients would begin palliative treatment as soon as radical treatment was considered to have failed. In this sense, the lack of evidence restricted the level of depth with which the progression of the disease could be represented within the model. These decisions were taken iteratively as we understood what evidence was available and were only formalised after we had begun to implement the model.

The model was implemented as a next-event discrete event simulation (DES) model. An individual-level simulation approach was taken as this allows for the more complex representation of model events conditional on patient characteristics and provides a greater level of flexibility in implementing and adapting the model as compared to a cohort

approach (e.g. a Markov model). The model was developed by considering the relevant competing events at each point in the clinical pathway (see Figure 3 and Figure 4). The time to each event was sampled for each patient, with the next event determined by whichever of these occurs first. After each event, an individual's risk profile is updated (for example, age and disease status) and the times to the next set of relevant events are re-calculated. Other-cause mortality is sampled differently in that this is defined upon model entry and the remaining time to this event is recalculated upon the occurrence of any other non-fatal event. Costs and effects are recorded as the patient progresses through the model, conditional on the events that they experience. A continuous discounting approach was adopted to account for health outcomes and costs which accrue over a particular time period. One-off costs (e.g. surgery) were discounted using a standard periodic discounting approach. The programming approach implemented within the final model followed the method suggested by Tappenden *et al.*⁵⁰

Detailed model description and programming logic

Patients enter the simulation model having been referred to secondary care by their GP, either due to the presence of symptoms or due to an elevated PSA test or from a secondary care physician who suspects the individual might have prostate cancer. The model design is summarised in Figure 3 and Figure 4 and the underlying logic of each section is described below. Each box within the diagram represents a SIMUL8 'workcentre' in which events, costs and consequences are sampled and applied to individual patients. With the exception of hormone and palliative treatments, all events which are modelled according to multiple competing risks are implemented using two related workcentres – one dummy workcentre which determines which event occurs next and one workcentre which represents the actual interaction of the patient with the prostate cancer service.

Workcentre 1: Initial characteristics

Upon entry into the model, patients are assigned initial characteristics. These include: the presence or absence of prostate cancer, age, initial stage (using standard Tumour Node Metastases [TNM] classification) and Gleason score. Patients are assigned a risk category based on the D'Amico classification using clinical stage and Gleason score (see Table 5), which later dictates which treatment options are available to the patient. PSA score is then sampled conditional on stage; this was necessary as the national registry data used to

assign patients' characteristics did not include data on initial PSA score (see Section "Evidence used to inform model parameters").

Published results from the observation arm of the Bill-Axelsson trial⁷⁷ were used to provide information on the natural history of prostate cancer for each disease event (local progression, metastases and prostate cancer death). Patients included in this arm of the trial were from an unscreened (Scandinavian) population and most did not receive any curative treatment.

The incidence of prostate cancer is not captured in the model - whether a man has prostate cancer or not is defined on entry to the model. If a simulated individual does not have prostate cancer on entry into the model, it is assumed that he cannot go on to develop prostate cancer. A proportion of these patients are assumed to have benign prostatic hyperplasia (BPH).

Figure 3. Design-oriented model: pre-clinical and diagnostic sections

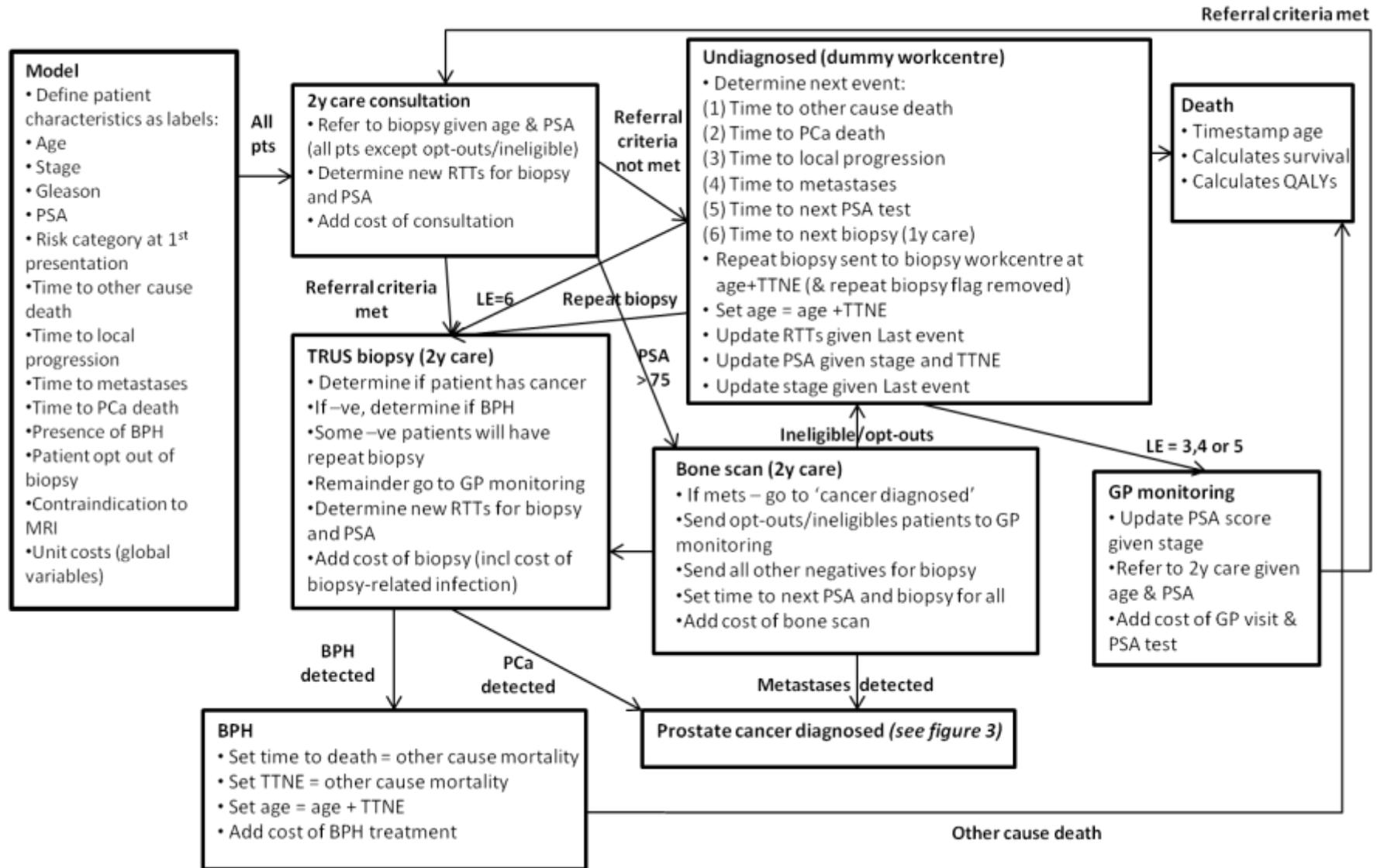
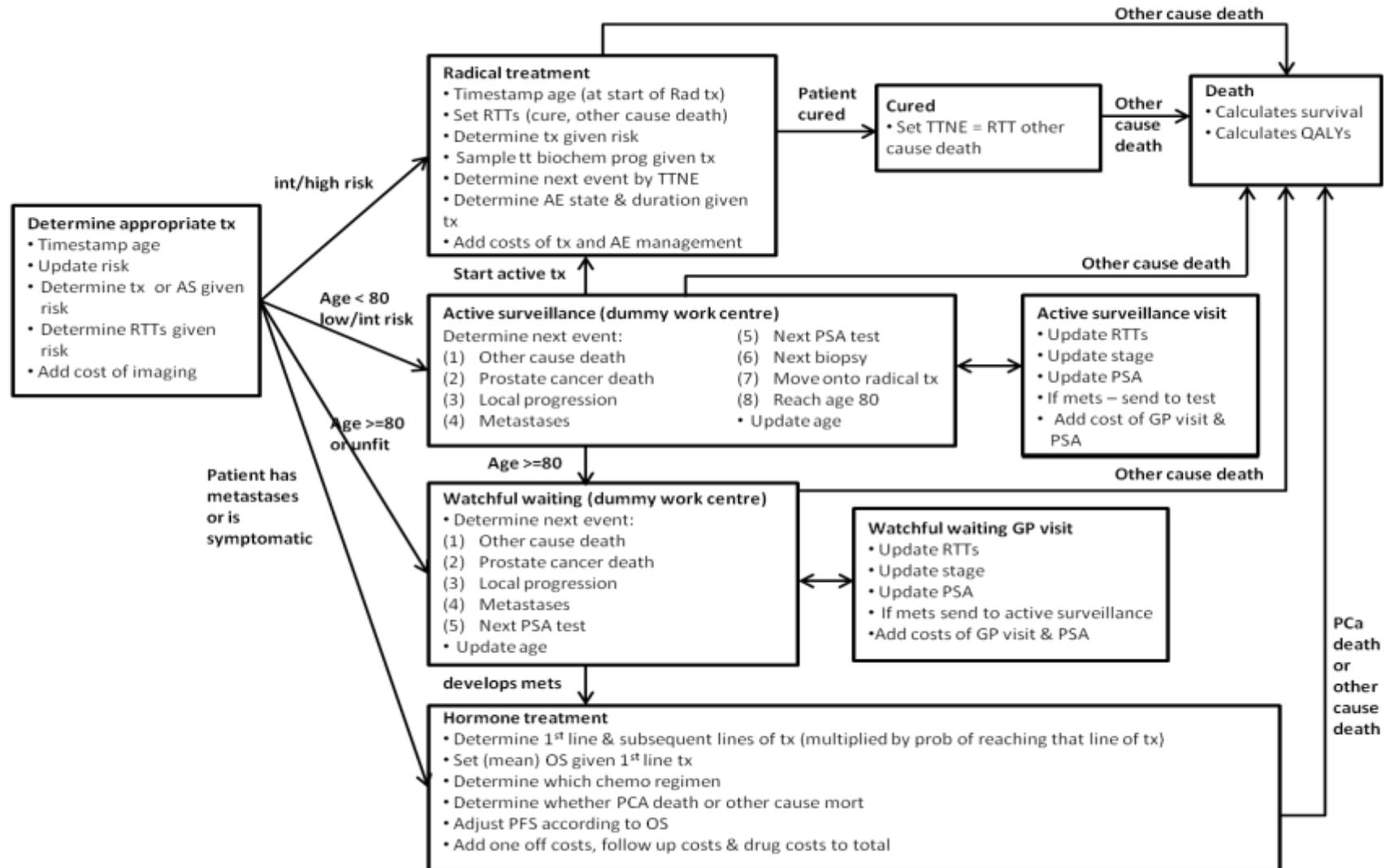


Figure 4. Design-oriented model: treatment pathways for diagnosed disease



Workcentre 2: Secondary care attendance

Following referral to secondary care, all patients are assumed to have a repeat PSA test (from a blood test) and DRE. Patients with a very high PSA score ($>75\text{ng/ml}$), which is taken to indicate obvious symptoms of advanced prostate cancer, are offered a bone scan without prior biopsy and hormone treatment with palliative intent. All other patients are considered for a TRUS biopsy if they meet the Prostate Cancer Risk Management Programme (PCRMP) primary care referral guidelines,⁷⁸ which are dependent on age and PSA score. Patients who do not meet the referral criteria, who have already had three prior biopsies or who opt out of biopsy, are sent for GP monitoring for 6-monthly PSA tests. These patients do not have a diagnosis of prostate cancer, although some may have underlying disease which may or may not be diagnosed if they re-enter secondary care. If a patient has undiagnosed prostate cancer the disease will progress untreated. Patients who do not have prostate cancer are assumed not to develop prostate cancer within their lifetime. For the sake of simplicity, it is assumed that no time elapses between the secondary care visit and the primary care attendance (either at model entry or when the patient attends GP monitoring). The cost of the PSA test is added, but contact between the patient and his GP is not included.

Workcentre 3: TRUS guided prostate needle biopsy

Upon entry to the biopsy workcentre, the number of biopsies is recorded as we assumed that patients could undergo a maximum of three biopsies in their lifetime, unless they are on active surveillance. The probability that a patient receives a positive biopsy result is based on the sensitivity of the test given the individual's true underlying histology. TRUS is assumed to be perfectly specific, meaning that all men who do not have prostate cancer will be correctly identified as not having the disease. The results of a TRUS biopsy given the presence/absence of underlying cancer are sampled and patients with a true positive result are sent to the 'determine appropriate treatment' workcentre. A proportion of patients who test negative are assumed to be invited to attend a repeat biopsy in 6 months' time, whilst the remainder are assumed to return to GP monitoring and undergo a PSA test in 6 months' time. Those patients who have undiagnosed BPH are assumed to have this pathology detected at this point and remain in the BPH workcentre until they die of other causes. Those patients who test negative, do not have BPH and were not referred for a repeat biopsy, undergo GP monitoring – these patients may have prostate cancer, but this is

clinically unknown at this stage. The model assumes that patients do not attend every GP visit to which they are invited. Where applicable, the cost of TRUS biopsy is added to the running total cost. A probability of experiencing infection due to TRUS is also sampled and the cost of treating the infection, if it occurs, is added to the running total.

Workcentre 4: Undiagnosed (dummy workcentre)

Patients who enter the GP monitoring workcentre do not yet have and may never receive a diagnosis of prostate cancer; these patients may or may not have underlying cancer. These patients are assumed to undergo PSA tests every 6 months indefinitely. For these patients, the time to the next event (denoted TTNE) is then determined. Competing events are: (1) other cause mortality, (2) prostate cancer-specific death, (3) local progression (unless this has already occurred), (4) metastases (unless this has already occurred), (5) next scheduled PSA test and (6) time to next biopsy (for those with a scheduled repeat biopsy only). If cancer-specific or other cause death occurs during GP monitoring, patients exit the model at this point. The remaining time to each competing event is then recalculated based on the time interval TTNE. If the next event is local progression or metastases, this is assumed to manifest symptomatically and triggers a GP visit and PSA test at the time of the clinical event. Most patients return to the GP for their scheduled 6-monthly PSA test (a proportion are assumed to not attend). If the patient was due to undergo a repeat biopsy but some other event occurs first, this is assumed to result in earlier biopsy (at $\text{age} + \text{TTNE}$). Age is then updated by the time to next event for all patients.

Workcentre 5: Primary care appointment for PSA test

In the primary care workcentre, patients who meet the PCRMP referral guidelines (dependent on age and current PSA score) are assumed to be sent for a biopsy. Those patients who do not meet the referral criteria, who have already had three prior biopsies or who opt out of biopsy return to GP monitoring with the next PSA test scheduled for 6 months after the current event. The cost of a primary care visit plus the cost of the PSA test is added to the running total. The time of the PSA test is recorded.

Workcentre 6: Bone scan

Bone scans are assumed to be perfectly sensitive and specific within the model; this is a simplifying assumption due to the lack of evidence. Patients who have metastases are

assumed to be identified by the scan; these patients are diagnosed at this point and go on for treatment planning. Patients who do not have metastases are correctly identified and if eligible will have a biopsy immediately or will otherwise have GP monitoring. The cost of the bone scan is added to the running total cost.

Workcentre 7: Determine appropriate treatment

At the point of diagnosis all patients enter this workcentre to determine appropriate treatment given their age, stage of disease and suitability for radical treatment. Patient risk, according to the D'Amico classification, is updated at this point based on the patient's current underlying cancer stage, Gleason score and PSA score. Disease stage is updated over time in line with the disease logic model detailed in Figure 2. PSA score is updated only when patients receive GP monitoring. Gleason score is not updated over time (note this assumption has been made elsewhere).⁷⁹

If the patient has metastases they are assumed to receive palliative hormone treatment. If the patient is below the age of 80, is suitable for radical treatment and has low-risk disease they are assumed to go to active surveillance with the intention of later receiving radical treatment, either at the onset of symptoms or when they choose to undergo treatment. If the patient is below the age of 80, is suitable for radical treatment and has intermediate-risk disease they are assumed to either transit immediately to radical treatment or to enter into active surveillance. High-risk patients are assumed to transit immediately to radical treatment. Patients who are unsuitable for radical treatment and are symptomatic are assumed to transit immediately to palliative hormone treatment. Patients who are unsuitable for radical treatment and are not symptomatic are assumed to receive watchful waiting. If the patient has metastases but has not previously had a bone scan since developing metastases they receive a bone scan at this point. All patients undergo an MRI scan or CT scan prior to receiving radical treatment.

Workcentre 8: Active surveillance (dummy workcentre)

This is a 'dummy' workcentre which determines the next relevant event for a given patient. As noted above, only patients with low- or intermediate-risk disease enter AS. Upon entry into AS, patients are assumed to undergo a PSA test every 3 months for the first year after their initial diagnosis of prostate cancer and every 6 months thereafter until they leave

surveillance or die. TRUS biopsy is assumed to take place one year following initial diagnosis and then every three years thereafter until they leave surveillance or die. Patients who experience local progression or those who opt for treatment over surveillance go on to receive radical treatment. Whilst the model assumes that it is impossible for patients to develop metastatic disease on active surveillance, we do not believe that this is a strong assumption as in reality metastasis is very unlikely to occur in these patients. Patients who reach the age of 80 without having radical treatment are assumed to transit to watchful waiting and are assumed to be no longer suitable for radical treatment.

Workcentre 9: Active surveillance visit

Patients enter the active surveillance visit workcentre if their last event was non-fatal. At this point individuals can either receive a scheduled test (PSA or biopsy) or receive radical treatment (determined by either patient choice or due to symptomatic disease progression). The model assumes that every patient undergoes a PSA test on entry into the workcentre. The last event is used to update the time to the next events in the model. For example if a patient reached 80 years of age and was moved onto watchful waiting, the time to the next PSA test is dictated by the watchful waiting test schedule rather than the AS test schedule.

Workcentre 10: Watchful waiting GP visit

Patients on watchful waiting are assumed to undergo a PSA test every 12 months indefinitely. The time to the next events are updated. Stage is updated in the model if the disease has progressed. If a patient developed metastatic disease, the model assumes that hormone treatment will be initiated. Otherwise, the patient will remain on watchful waiting. The cost of the scheduled GP consultation and the PSA test is added to running total of costs.

Workcentre 11: Watchful waiting (dummy workcentre)

This dummy workcentre calculates the event occurs next. This can be either other cause death, disease progression (local or metastatic) or a scheduled test. Disease progression is assumed to be symptomatic so a patient will present outside their scheduled appointment and will be offered hormone treatment.

Workcentre 12: Radical treatment

Patients who enter the radical treatment workcentre do not have metastatic disease and are classified according to the D'Amico risk criteria at diagnosis.⁸⁰ All low-risk patients will have previously been on AS, but have switched onto radical treatment (thus their real risk status at this point may no longer be considered low). The model assumes that these 'low-risk' patients are offered the same treatment as patients with intermediate-risk disease: radical prostatectomy (open), radiotherapy (and hormones) or brachytherapy. High-risk patients or patients with locally advanced disease are only eligible for hormones plus radiotherapy or hormone therapy alone.

Radical treatment is assumed to have an impact on time to local progression and the frequency of three adverse events (sexual dysfunction, urinary dysfunction and bowel dysfunction). The model assumes that time to prostate cancer death is not directly influenced by radical treatment. Patients having radical prostatectomy may die peri-operatively due to surgical complications, with risk increasing with age. As noted earlier, the model equates biochemical progression (the primary outcome from trials of radical treatment) with local disease progression.

The time to local progression and the frequency of three main adverse events are defined according to the patient's allocated treatment option. The three adverse events included in the model are associated with different dis-utilities, which we assume to be life-long and additive (that is the impact on health-related quality of life for each adverse event is independent of other adverse events). The model assumes that outcomes from treatment are the same for all risk categories since the available RCT evidence does not suggest otherwise.

If patients do not die of other causes, they will receive follow up comprised of an annual bone scan, a PSA test every 6 months and a urology consultation. In the first two years following treatment the PSA test will be done in secondary care and the consultation as an outpatient visit. After that, the PSA test will be undertaken in primary care and the consultation will be by telephone with a urology consultant. Follow up is assumed to cease at the time of local progression (or death from other causes).

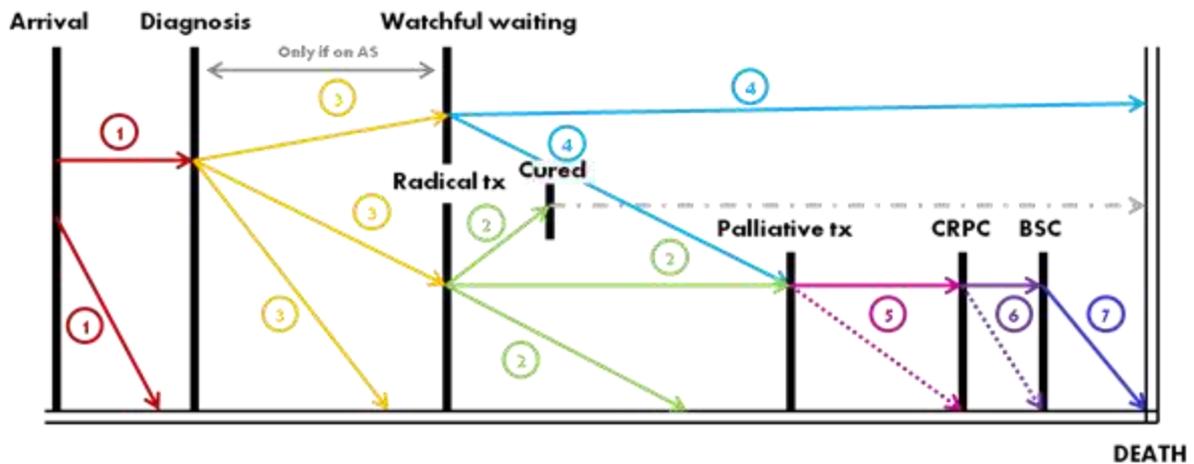
Workcentre 13: Hormone treatment plus chemotherapy plus best supportive care

The hormone treatment plus chemotherapy plus best supportive care (BSC) workcentre calculates the patient's time to prostate cancer death and determines the proportion of this period which is 'progression-free'; this is assumed to be dependent on the treatment received. The remaining time to other-cause death remains unaffected by treatment. The model assumes that first-line treatment (intermittent hormones, continuous hormones, bilateral orchidectomy or bicalutamide monotherapy) determines overall survival and the sequence of later lines of treatment. Progression-free survival (PFS) from each line of treatment (up to four lines of treatment in the base case analysis) is added, with any remaining time before prostate cancer death spent in a progressive disease state whilst receiving best supportive care. If patients survive the first three lines of treatment, chemotherapy is given as the fourth-line treatment: either using a docetaxel-based or mitoxantrone-based combination regimen. A fixed proportion of patients will not receive chemotherapy (not all patients will be fit enough). Due to evidence limitations, mean health state sojourn times are used, so all patients allocated to the same treatment will have the same outcomes (except for other-cause death). Cause of death is determined (prostate cancer or other cause) and PFS is adjusted to ensure the sum of progression-free intervals does not exceed the assigned overall survival. That is, if the sum of progression-free intervals exceeds the overall survival for an individual patient, the final progression-free survival interval is truncated.

Workcentre 14: Death

Health outcomes are calculated for each simulated patient at time of death. The cost of terminal care is added here, if the patient has died from prostate cancer. Survival is calculated by adding together the time each patient has spent in different segments of the model. There are up to seven time segments which reflect all possible paths through the model, as shown in Figure 5.

Figure 5. Time segments (numbered) used to calculate overall survival



The numbered time segments in Figure 5 refer to the following routes through the model:

- Segment 1: Initial attendance to death (non-cancer or undiagnosed cancer patients) or cancer diagnosis.
- Segment 2: From the start of radical treatment to cure, biochemical relapse or other-cause death.
- Segment 3: From the start of active surveillance to initiating radical treatment, until death or until beginning watchful waiting (note: if patients do not receive active surveillance no time will be spent in this segment).
- Segment 4: From the start of watchful waiting to the start of palliative treatment, or death.
- Segment 5: Hormone treatment to end of progression free survival from third-line (palliative) treatment (i.e. men with Castration Refractory Prostate Cancer (CRPC)).
- Segment 6: From the start of fourth-line (palliative) treatment to beginning of Best Supportive Care (BSC) or death.
- Segment 7: BSC to death.

Discounted and undiscounted life years and QALYs are calculated for all patients. The lifetime costs of adverse events are added in the death workcentre, including the cost of

screening with flexible sigmoidoscopy every five years for patients who receive radical radiotherapy, in line with the CG58 recommendation.⁵⁵

Evidence used to inform the model parameters

The model was populated using evidence identified within the 2008 NICE Prostate Cancer Guideline⁷¹ supplemented with additional evidence identified through rapid literature searches and/or expert opinion. We did not conduct systematic reviews for all of these parameters, as this was not possible within the resources available for the study, and there are certain parameters (e.g. unit costs) whereby a conventional systematic review approach is neither required nor preferred.⁵⁷ This is likely to mirror the pragmatic approach taken to populate health economic models during routine development of NICE clinical guidelines. The model includes the following groups of parameters:

- Disease epidemiology and baseline patient characteristics: incidence and prevalence of the condition and subgroups, baseline risks, rates of progression of disease and mortality rates.
- Test operating characteristics (e.g. sensitivity and specificity) of the tests included in the pathway.
- Clinical effectiveness of the treatments included in the pathway (e.g. progression-free survival, time to biochemical relapse, peri-operative mortality.).
- Patient behaviour (e.g. probabilities of opting out of biopsies), probability of attending routine PSA tests.
- Utilities associated with disease, treatment and adverse events.
- Resource use and unit costs.

Disease epidemiology and baseline patient characteristics

Data were required to define the initial characteristics of men with and without prostate cancer. We used national cancer registry data obtained from the South West Public Health Observatory (SWPHO) to provide information on age, clinical stage at diagnosis and Gleason score at diagnosis for patients with diagnosed prostate cancer (South West Public Health Observatory 2010, data held on file). The national registry database does not record PSA score, hence it was necessary to calculate patients' risk according to two of the three D'Amico criteria (see Table 5). Age-specific PSA values from men in the watchful waiting arm

of the Bill-Axelson RCT⁷⁷ were used to estimate PSA scores on model entry for men with prostate cancer (cited in Tilling *et al*⁸¹).

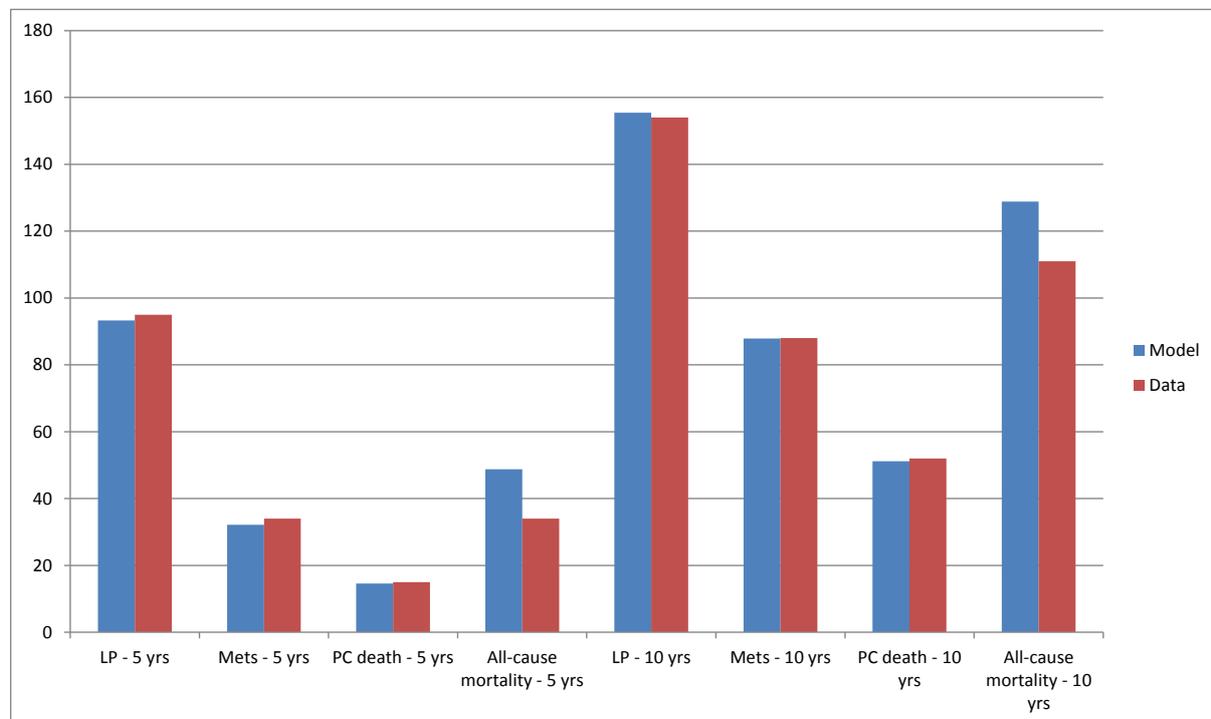
Evidence relating to age-specific PSA scores of the cancer-free population in the model was taken from the Krimpen *et al* longitudinal community-based study.⁸² We have no UK data relating to this patient group so we assumed that PSA scores in these patients follow the same age distribution as for men with prostate cancer.

There is uncertainty regarding the true disease prevalence in men referred to secondary care with suspected prostate cancer. Owing to an absence of empirical estimates, we assumed a value of 25% based on expert opinion which roughly reflects the results from non-UK autopsy studies (20-34%).⁸³⁻⁸⁵ Data on death from causes other than prostate cancer were taken from national life tables; these were adjusted by removing all deaths attributed to prostate cancer.⁸⁶

Independent survival curves for local disease progression, metastatic disease progression and prostate cancer death were taken from the 2011 publication of the Bill-Axelson *et al* RCT⁷⁷ which compared radical prostatectomy with watchful waiting; this was the closest proxy to information on the natural history of the disease without treatment. This study reported numbers of patients who experienced local progression, metastases and prostate cancer death at 5-year and 10-year time points. It should be noted that the Bill-Axelson trial outcomes relate to the point of documented progression and metastases rather than the true underlying time of histological change. These outcomes are also based on a Scandinavian population of men in the pre-PSA testing/screening era hence they may not fully reflect the UK population within the model. We used model calibration methods to derive correlated conditional distributions for these events. We implemented a random-walk variant of the Metropolis-Hastings algorithm⁸⁷ based on the methods described by Whyte *et al*⁸⁸ directly in SIMUL8 and fitted the model against the unconditional data from Bill-Axelson *et al* and other-cause mortality estimates from the UK. We ran the algorithm over four separate chains with different starting vectors in order to estimate plausible distributions for each event, conditional on the population having experienced the previous event. The joint distributions of progression parameters were used directly in the probabilistic sensitivity analysis. Figure 6 presents a comparison of the maximum *a posteriori* estimates produced by the calibration process against the observed data reported

by Bill-Axelson *et al*; the figures show that the calibration provides a good fit to the observed data.

Figure 6. Observed vs. calibrated frequencies of disease events



LP=local progression; Mets=metastases; PC=prostate cancer

Diagnostic test accuracy

We assumed a sensitivity of 77% for TRUS biopsy.⁸⁹ We assumed that PSA, DRE, MRI, CT and bone scans are perfect tests. We also assumed that the TRUS guided biopsy is perfectly specific (i.e. no false-positive results) whereas in reality TRUS may lead to over-treatment. Test accuracy studies are difficult to undertake in this area, since pathological confirmation will not be carried out for patients with negative biopsy results. The simplifying assumptions were necessary not only because of the lack of gold-standard comparison studies but also due to the complexity of including the implications of misdiagnosis and misclassification from these tests in the model and the limited information available on the natural history progression of prostate cancer.

A small proportion of patients will experience an infection as a result of biopsy, (probability=0.47%) and this is represented in the model.⁹⁰ Not all patients are willing to

undergo biopsy; we assume 12% men will opt out.⁹¹ Uncertainty surrounding these parameters was characterised using beta distributions.

Clinical effectiveness

Where more than one treatment is recommended at a particular point in the pathway, we used proportions elicited from the Department of Health National Radiotherapy Group and experts on the NICE Prostate Cancer Guideline Development Group. The management and treatment options in the model were grouped according to their clinical intent (e.g. delaying and/or avoiding recurrence or increasing progression-free survival) and the key outcome measures used in the clinical studies from which efficacy estimates were drawn. Perhaps surprisingly, there is a lack of evidence on the comparative effectiveness of currently available radical treatments for prostate cancer. Therefore, through necessity, evidence from different trials was used and compared against single arms of other trials using naïve indirect comparison methods (see Table 6). Radical prostatectomy is also associated with an excess mortality risk.⁹² As discussed above, biochemical relapse after radical treatment is used as a proxy for local progression due to a lack of direct evidence on local progression *per se*. The estimates used in the model, characterised in terms of first- and second-order uncertainty, are detailed in Table 6.

Palliative treatments (Table 7) were also difficult to model as we did not identify any RCTs which explicitly evaluated planned sequences of treatments. Therefore, we assumed that first-line palliative treatment is the sole determinant of overall survival due to prostate cancer. Subsequent lines of treatment are assumed only to increase the proportion of the patient's remaining survival time that is progression-free. This manipulation of the evidence requires that we ignore first-order uncertainty in these parameters and therefore use mean sojourn times for estimates of overall and progression-free survival, which is not ideal. The uncertainty in these mean values is still however reflected in the probabilistic sensitivity analysis.

Table 6: Radical treatment efficacy and adverse event parameters

Treatment	Model parameter	First-order uncertainty	Second order uncertainty	Source
Radical prostatectomy	Time to local progression	Exponential ($\alpha=0.016$)	Normal ($\lambda= 0.016, se=0.002$)	Bill-Axelsson 2011 (radical prostatectomy vs. observation). Local progression at 15 years. ⁷⁷
	Probability of sexual dysfunction	Uniform (0,1)	Beta ($\alpha =168, \beta=121, mean = 0.58$)	
	Probability of urinary dysfunction	Uniform (0,1)	Beta ($\alpha =99, \beta=190, mean = 0.34$)	
	Probability of bowel dysfunction	0	0	
Brachytherapy	Time to local progression	weibull ($\alpha=0.846112974, \beta=2.80697845$)	Multivariate normal ($\log \lambda= -3.83, \gamma= 0.85$)	Giberti 2009 (radical prostatectomy vs. brachytherapy)
	Probability of sexual dysfunction	Uniform (0,1)	Beta ($\alpha =42, \beta=58, mean = 0.42$)	
	Probability of urinary dysfunction	Uniform (0,1)	Beta ($\alpha =80, \beta=20, mean = 0.8$)	
	Probability of bowel dysfunction	Not reported. In base-case analysis set equal to probability of bowel AE with radiotherapy.		Assumption based on Fransson 2009 (QoL data from SPCG7, Widmark RCT) ^{93;94}
Adjuvant hormones +radical radiotherapy	Time to local progression	weibull ($\alpha=1.354431605, \beta=21.78254729$)	Multivariate normal ($\log \lambda= -4.17, \gamma= 1.35$)	Widmark 2009 (adjuvant hormones + RT vs. hormones alone) ⁹⁴
	Probability of sexual dysfunction	Uniform (0,1)	Beta ($\alpha =250, \beta=85, mean = 0.75$)	
	Probability of urinary dysfunction	Uniform (0,1)	Beta ($\alpha =64, \beta=289, mean = 0.18$)	
	Probability of bowel dysfunction	Uniform (0,1)	Beta ($\alpha =37, \beta=312, mean = 0.1$)	
Hormone therapy alone	Time to biochemical progression	weibull ($\alpha=1.06, \beta=5.57$)	Multivariate normal ($\log \lambda= -1.82, \gamma= 1.06$)	Widmark 2009 ⁹⁴
	Probability of sexual dysfunction	Uniform (0,1)	Beta ($\alpha =197, \beta=110, mean = 0.64$)	Fransson 2009 ⁹³
	Probability of urinary dysfunction	Uniform (0,1)	Beta ($\alpha =39, \beta=289, mean = 0.12$)	
	Probability of bowel dysfunction	Uniform (0,1)	Beta ($\alpha =23, \beta=312, mean = 0.07$)	

Table 7: Palliative treatment efficacy parameters

Treatment	Progression-free		Overall survival		Source	Comments
	Mean (years)	Second order uncertainty	Mean (years)	Second order uncertainty		
First-line: Intermittent hormones	7.4	Multivariate normal (log λ = -2.43, γ = 1.18)	7.0	Multivariate normal (log λ = -2.81, γ = 1.38)	Calais da Silva ⁹⁵	
First-line: Continuous hormones	13.5	Multivariate normal (log λ = -2.37, γ = 0.92)	7.2	Multivariate normal (log λ = -2.22, γ = 1.11)	Calais da Silva ⁹⁵	Same data used in the model when given as second-line treatment (when patient has received first-line bicalutamide monotherapy).
First-line: Bilateral orchidectomy	3.6	Multivariate normal (log λ = -1.23, γ = 0.99)	3.4	Multivariate normal (log λ = -2.05, γ = 1.54)	PFS: Eisenberger ⁹⁶ OS: Seidenfeld ⁹⁷	
First-line: Bicalutamide monotherapy	1.2	Multivariate normal (log λ = -0.52, γ = 1.61)	2.8	Log normal (ln(mean)=0.18, se=0.11)	Tyrrell ⁹⁸	Hazard ratio applied to bilateral orchidectomy baseline.
Second-line: LHRHa + Bicalutamide	0.5	Normal (mean = 5.8 months, sd=0.2948)	n/a	n/a	Suzuki ⁹⁹	2nd line CAB, but pts have had 1st line CAB (no pts in our model have had this intervention). Note patients in model will not receive this intervention if they have previously had bicalutamide monotherapy.
Third-line: LHRHa + dexamethasone	0.8	Multivariate normal (log λ =0.11, γ = 1.23)	n/a	n/a	Venkitaraman ¹⁰⁰	PFS curve was supplied by author on request 10/12/11.
Fourth-line: Docetaxel + prednisolone	0.7	Multivariate normal (log λ =0.35, γ = 1.31)	n/a	n/a	Petrylak ¹⁰¹	TAX327 used in NICE TA101 (Tannock, 2004) was not used as PFS was not measured in the trial. Regimen in Petrylak was Docetaxel +estramustine, which is assumed to be equivalent to Docetaxel + prednisolone.
Fourth-line: Mitoxantrone + prednisolone	1.4	Multivariate normal (log λ =1.07, γ = 0.54)	n/a	n/a	Petrylak{6057}	

Health utilities

The lack of published evidence relating to the impact of prostate cancer and its treatment on health-related quality of life has been widely acknowledged. The health utility values used in the model were drawn from recent economic evaluations of prostate cancer (see **Error! Reference source not found.**). We did not identify any HRQoL evidence published after these studies.

Table 8: Utility data

Treatment	Mean utility value	Second order uncertainty	Source
Active surveillance	0.81	1- beta (4, 0.0675)	Hummel ⁹⁹
Radical treatment	0.78	1- beta (4, 0.055)	Hummel ⁹⁹
Local progression	0.73	1- beta (4, 0.08)	Hummel ⁹⁹
Hormone-refractory prostate cancer	0.64	1- beta (4, 0.14)	Hummel ⁹⁹

We incorporated the HRQoL impact of the three most common adverse events attributable to radical treatment; bowel function, urinary function and sexual function as disutilities (see table 9). Owing to the absence of data on the duration of adverse events, the model assumes that these last for the remaining lifetime of the patient. The impact of this assumption is not tested further here, however the flexibility of the model allows such assumptions to be amended easily. Owing to a lack of identified evidence, the differential impact of adverse events on health utilities due to specific palliative treatments was not captured.

Table 9: Disutility data

Treatment	Mean disutility value	Second order uncertainty	Source
Sexual dysfunction	0.10	1- beta (2.6, 23.40)	Krahn <i>et al.</i> cited in Chilcott 2010 ⁶⁹
Urinary dysfunction	0.06	1- beta (2.76, 43.24)	Krahn <i>et al.</i> cited in Chilcott 2010 ⁶⁹
Bowel dysfunction	0.11	beta (53.46, 6.61)	Krahn <i>et al.</i> cited in Chilcott 2010 ⁶⁹
Progression with hormone-refractory	0.07	1- beta as integer (7,93)	NICE TA255, ERG report

Resource use and unit costs

In accordance with the perspective of this analysis, the only costs considered were those relevant to the UK NHS & PSS. Costs were estimated in 2010-11 prices. Resource use estimates were drawn from the NICE Prostate Cancer Guideline CG58 recommendations⁷⁰ and were augmented with expert opinion (see Appendix 4). The cost of primary care contact before initial referral, primary care services during prostate cancer treatment and cardiovascular screening were not included in the model as these were difficult to ascertain from the current guideline recommendations and resource use patterns are likely to vary. In order to reflect the additional terminal care costs incurred by patients in the last month of life, a one-off cost of just over £4,000 was applied to men who died of prostate cancer. This cost was used in the NICE Technology Appraisal TA101 having been estimated from costing data originally supplied by Sanofi-Aventis on men with hormone-refractory disease.¹⁰³

Table 9: Resource use parameters

Treatment	Mean unit cost	SE (estimated)	Distribution	Source – NHS Reference Costs ¹⁰⁴ unless otherwise stated
PSA test in primary care	£11	£2.2	gamma	Hummel ¹⁰²
PSA test in secondary care	Assumed to be the same as PSA in primary care (above)			
Digital rectal examination	0	will be carried out as part of the consultation with the urologist		
Transrectal ultrasound guided biopsy	£200	£5.3	normal	HRG code LB27Z.
CT scan	£100	£2.4	normal	HRG code RA08Z.
MRI scan	£218	£6.0	normal	HRG code RA01Z.
Bone scan	£181	£5.5	normal	HRG code RA36Z.
Flexible sigmoidoscopy	£219	£8.8	normal	HRG code FZ54Z.
Appointment with GP (incl. training)	£36		fixed	PSSRU ¹⁰⁵
Appointment with GP practice nurse (incl. training)	£13		fixed	PSSRU ¹⁰⁵
Face-to-face consultation with Urology consultant (first)	£130	£3.1	normal	Outpatient attendance.
Face-to-face consultation with Urology consultant (follow up)	£91	£1.4	normal	Outpatient attendance.
Face-to-face consultation with	£148	£3.2	normal	Outpatient

Treatment	Mean unit cost	SE (estimated)	Distribution	Source – NHS Reference Costs ¹⁰⁴ unless otherwise stated
Surgical consultant (first)				attendance.
Face-to-face consultation with Surgical consultant (follow up)	£106	£2.4	normal	Outpatient attendance.
Face-to-face consultation with Clinical oncology consultant (first)	£180	£4.6	normal	Outpatient attendance.
Face-to-face consultation with Clinical oncology consultant (follow up)	£122	£5.6	normal	Outpatient attendance.
Face-to-face consultation with Medical oncology consultant (first)	£171	£5.6	normal	Outpatient attendance.
Face-to-face consultation with Medical oncology consultant (follow up)	£120	£4.2	normal	Outpatient attendance.
Telephone follow up with urology consultant	£54	£6.2	normal	Urology consultant led follow up non face-to-face.
Oral administration of chemotherapy (first)	£171	£7.6	normal	HRG code SB11Z.
Parenteral administration of chemotherapy (first)	£265	£8.5	normal	HRG code SB13Z.
Administration of subsequent elements of a Chemotherapy cycle	£294	£9.1	normal	HRG code SB15Z.
Radical prostatectomy	£5,119	£128.4	normal	HRG code LB21Z
Bilateral orchidectomy	£407	£22.2	normal	HRG code LB35B
Conformal radiotherapy planning	£581	£81.0	normal	HRG code SC51Z
Delivery of conformal radiotherapy	£111	£6.2	normal	HRG code SC23Z
Delivery of external beam radiotherapy	£91	£4.1	normal	HRG code SC23Z
Brachytherapy planning	£1,123	£97.1	normal	HRG code SC54Z
Delivery of brachytherapy	£383	£196.6	normal	HRG code SC26Z
Specialist erectile dysfunction services	£179	£13.9	normal	HRG code LB43Z
Incontinence containments	£68	£33.8	gamma	Hummel ¹⁰²
Post biopsy infection requiring hospitalisation	£2,623	£79.9	normal	HRG code PA16B
Strontium (one dose)	£1,070	£0.55	normal	HRG code SC29Z
Transurethral resection of the prostate (ablative procedure for BPH)	£2,405	£37.8	normal	HRG code LB25B
Terminal care	£4,142		fixed	Collins 2007 ¹⁰³ , inflated to 2010-11 prices

Drug costs used in the base case analysis were based on prices listed in the British National Formulary, as shown in table 6.¹⁰⁶

Table 6 BNF drug costs

Drug	List price, BNF ¹⁰⁶
Docetaxel	£7,487
Prednisolone	£227
Mitoxantrone	£914
Cyproterone acetate	£823
Stilboestrol	£2,150
Bicalutamide 150mg	£127
Bicalutamide 50mg	£57
Flutamide	£491
Dexamethasone	£1,983
Sildenafil	£1,005

Handling uncertainty

With the exception of PSA trajectories which are sampled only according to first-order uncertainty, the model is fully probabilistic. Sampling of parameter uncertainty for the probabilistic sensitivity analysis was implemented by sampling the necessary distributions externally in Microsoft Excel[®] (Microsoft Corporation, Redmond) and reading them into SIMUL8. This approach has the added advantage that changes in model results reflect only the impact of changes to the pathway (e.g. new chemotherapy B vs current chemotherapy A) rather than randomness in the sampling of the parameters that make up the model structure; a similar approach was also used in the second case study (see Chapter 5). One thousand five hundred probabilistic samples were used to propagate parameter uncertainty through the model, and all headline results are presented as the expectation of the mean rather than point estimates of parameters.

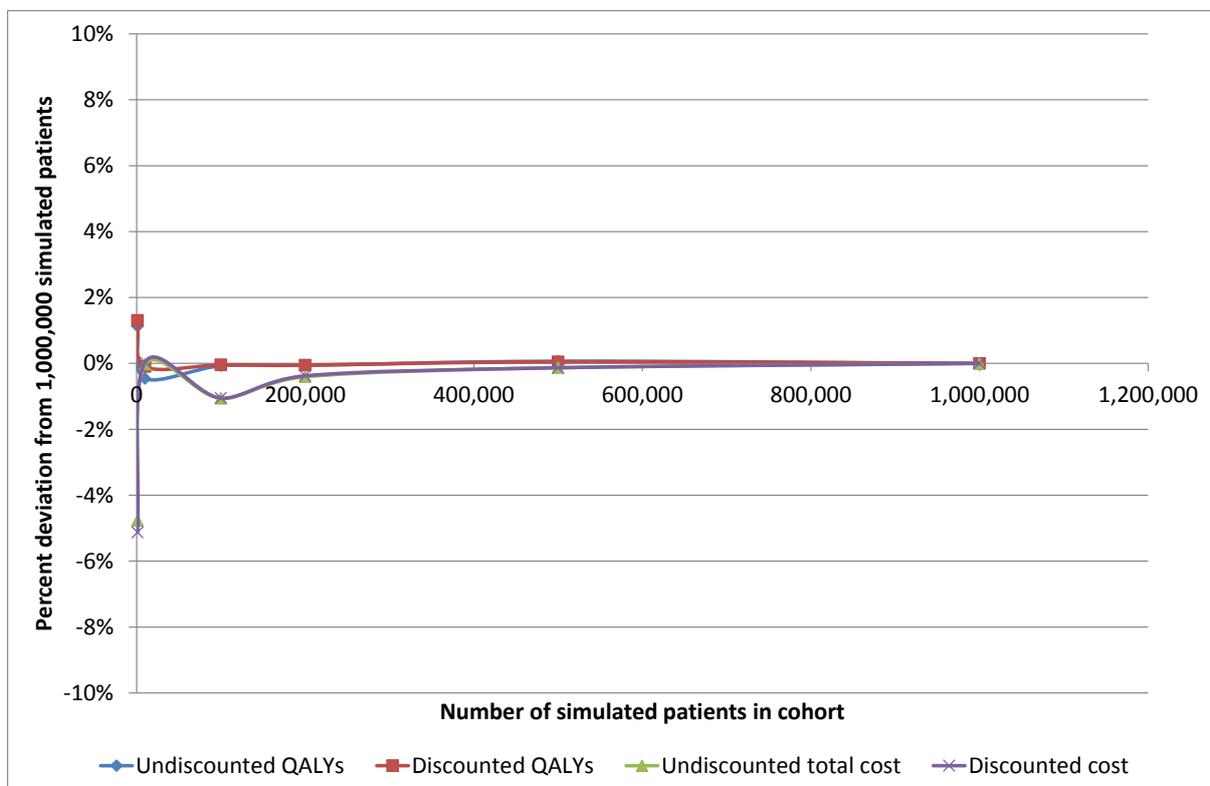
Verification and validation

Errors and inconsistencies in the model were checked for throughout the model development process, following the methods set out by Chilcott *et al.*²⁵ The model was verified internally (to ensure correct programming) and validated externally (to ensure consistency with expected results – for example, that survival times and levels of service use

are realistic). A variety of methods were used including black box testing (testing the behaviour of the model) and white box testing (scrutinising the programming code). In addition, the model was programmed to record intermediate model outcomes (e.g. survival contributions attributable to particular segments of the pathway and costs associated with specific workcentre) in order to provide a basis for whether changes to the pathway impacted upon those parts of the model as expected.

Once we were satisfied that the model was behaving as intended, we then assessed the number of patients required to achieve stability in the model results. We adopted a pragmatic approach to this using the results of the base case model only. We ran the model with the base case service configuration with different numbers of patients and compared the results from each section to the results for 1,000,000 simulated men. Figure 7 indicates that the costs and QALYs become fairly stable (<2% deviation) at around 100,000 simulated patients. Conservatively, we adopted a cohort of 200,000 simulated individuals for the economic analysis.

Figure 7. Model stability according to size of simulated cohort

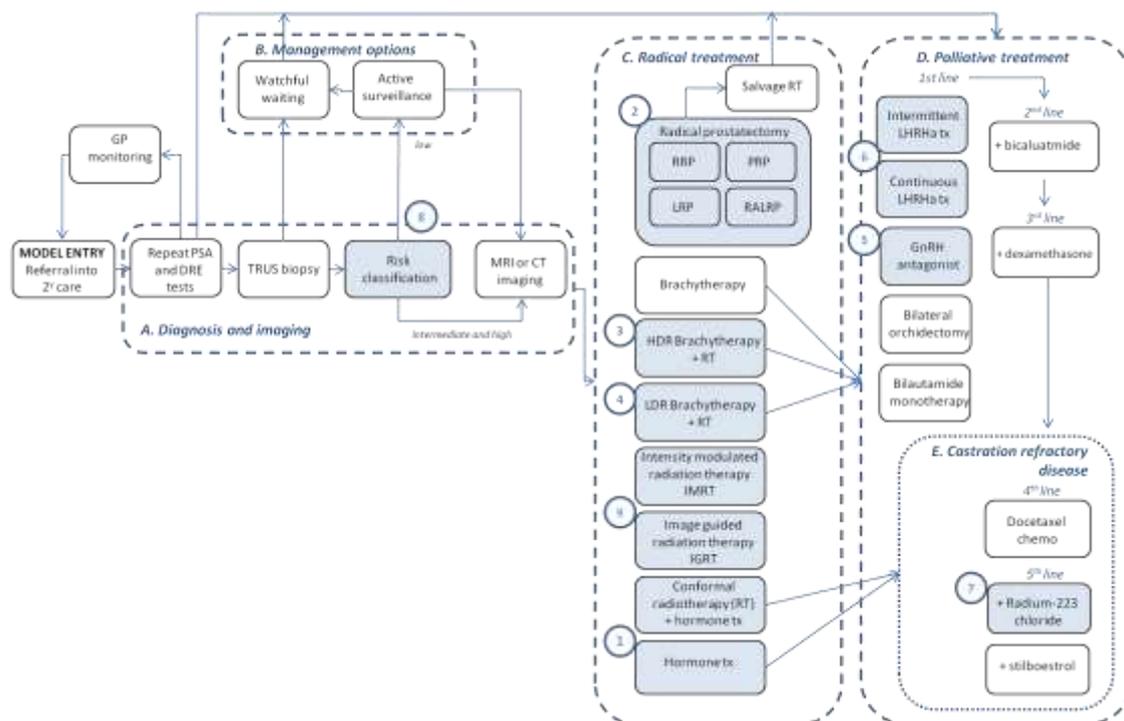


Modelling decision options across the service pathway

Nine topics were shortlisted from topics highlighted by NICE for possible inclusion in an update of the 2008 prostate cancer guideline (see Box 7 in chapter 3). Details of how the nine topics were shortlisted are discussed in Chapter 3.

Each topic implied an alternative clinical pathway, incorporating one or more changes to the recommendations made in CG56.⁵⁵ Figure 8 shows where these alternative recommendations are located in the clinical pathway. Each topic was transformed into a PICO-style review question, described below.

Figure 8. Location of topics (numbered) in the clinical pathway



Topic A - Pelvic radiotherapy with adjuvant hormone therapy for men with localised prostate cancer

The stated patient population for this topic – men with localised prostate cancer - is broad but reference is made to the SPCG-7 trial⁷⁷ which included men with locally advanced or high risk localised prostate cancer. The NICE Prostate Cancer Guideline⁵⁵ recommended that these patients should be offered either radiotherapy with hormone treatment or hormone treatment alone. In practice, many men will only be offered hormone treatment, without the option of additional radiotherapy. A focussed literature

search conducted by NICE¹⁰⁷ identified three published papers from two new RCTs.^{93;108} Only one paper had published full results of the trial at the time of analysis (the SPCG-7 trial⁹⁴). Six RCTs identified in the NICE Prostate Cancer Guideline⁷¹ have published additional follow up results with findings in support of combined radiotherapy and hormone therapy. Additionally, two observational studies that compared quality of life following radiotherapy plus hormone therapy to that following radiotherapy alone were identified.

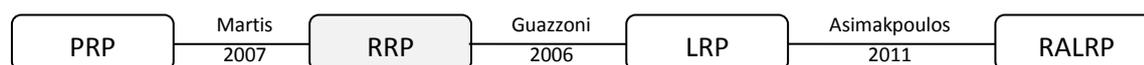
The PICO question was formulated to mimic the clinical question addressed in the only new RCT published in full since 2008, the SPCG-7 trial. Combined hormone treatment plus radiotherapy versus hormone treatment alone for men with locally advanced or high risk localised prostate cancer. Since the SPCG-7 trial was used to populate the base case model this economic question was evaluated without needing to modify the model structure.

Topic B - Surgical techniques for localised prostate cancer: open radical retropubic prostatectomy (RRP), trans-perineal prostatectomy (PRP), laparoscopic prostatectomy (LRP) or robot assisted laparoscopic prostatectomy (RALRP)

This topic suggests four alternative surgical techniques RRP, PRP, LRP and RALRP for men with localised prostate cancer undergoing surgery. NICE Prostate Cancer Guideline⁵⁵ did not recommend a specific procedure for radical prostatectomy. The base case model was populated with data from Bill-Axelsson *et al.* Accordingly, the cost used was the NHS reference cost for a standard open procedure.

Eleven studies were identified by NICE, including three systematic reviews of observational studies, three additional observational studies and three RCTs. The three RCTs investigate different pairwise comparisons, as shown in Figure 9.

Figure 9. RCT evidence network of surgical techniques for localised prostate cancer



We limited our analysis to RCT evidence only. The systematic reviews suggested some problems with the reliability of the observational evidence and in some cases the methods of synthesis do not appear to be robust. Of the three RCTs, only the trial reported by Martis

*et al*¹⁰⁹ provided longer term outcome data (biochemical recurrence); the others focussed on peri-operative outcomes.^{110;111} The time to biochemical recurrence survival curves reported for PRP and RRP in Martis *et al* are almost identical, hence the same curve that was used in the base case analysis was used for both procedures. It also seems generally reasonable, given the absence of evidence to suggest otherwise, that LRP and RALRP are also associated with the same biochemical recurrence rate.

Differential peri-operative mortality outcomes associated with specific techniques are not captured within the model. However differences in the frequency of adverse events associated with each surgical procedure are captured; RALRP is associated with fewer sexual and urinary problems than LRP, which has a similar adverse event profile to RRP and PRP (although LRP results in slightly more urinary problems than RRP and PRP).¹⁰⁹⁻¹¹¹

Another notable difference between the procedures is the difference in length of hospital stay for LRP or RALRP. We account for a shorter hospital stay for PRP (one third less than for RRP) as reported in Martis *et al*.¹⁰⁹ Robotic-assisted surgery (RALRP) is also associated with a shorter length of stay, estimated to be one day less than for the standard open procedure in a recent UK business case analysis (Oxford Radcliffe Hospitals NHS Trust). RALRP requires a significant capital investment which we include as an approximate figure of £3,000 per patient in addition to the non-capital costs, based on the most expensive robot system and assuming a fairly large centre with a throughput of around 150 patients per year (Ramsay *et al*, 2012).¹¹² RALRP is also associated with more costly consumables and a longer operating time (Oxford Radcliffe Hospitals NHS Trust). The current NHS reference cost for prostatectomy includes both RRP and PRP procedures. Laparoscopic prostatectomy is costed separately, but bundled with other laparoscopic urological procedures. Full details are given in appendix 5.

Topic C - High dose rate brachytherapy +external-beam radiotherapy for men with localised or locally advanced prostate cancer

Brachytherapy alone is currently recommended by NICE as an option for the treatment of men with intermediate- or low-risk disease but is not recommended for patients with high-risk disease. Brachytherapy combination therapy was not considered. The patient population overlaps with Topic D, hence we evaluated Topic C and D together.

Seven papers investigating HDR brachytherapy in combination with external beam radiotherapy were identified in a focussed search conducted by NICE; two of these were RCTs^{113;114} and five were observational studies.¹¹⁵⁻¹¹⁹ We restricted our analysis to use only the RCT data on biochemical relapse and frequency of adverse events.

Topic D - Low dose rate brachytherapy + external-beam radiotherapy for men with localised or locally advanced prostate cancer

LDR brachytherapy combination therapy was not considered in the NICE prostate cancer guideline.⁵⁵

No comparative data on the clinical effectiveness of LDR brachytherapy and external beam radiotherapy was identified. One US cohort study, reported by Sylvester *et al*,¹²⁰ was identified. This study reported 15-year follow up of 223 patients given I¹²⁵ or Pd¹⁰³ brachytherapy plus neoadjuvant radiotherapy. These data were used to estimate biochemical relapse-free survival and the frequency of adverse events.

Topic E – Degarelix (an LHRH antagonist) for men with locally advanced or metastatic prostate cancer

No luteinizing hormone releasing hormone (LHRH) antagonists were recommended in the NICE Prostate Cancer Guideline.⁵⁵ Degarelix is now licensed in the UK and was recently recommended by the Scottish Medicines Committee under a patient access scheme. One non-inferiority RCT comparing low dose degarelix (240/80 mg), high-dose degarelix (240/160 mg) and standard 7.5 mg monthly dose of leuprolide was identified.¹²¹ The primary outcome measure in this study was the cumulative probability of testosterone with other outcomes including the incidence of PSA failure. As this study only had 12-month follow up data and did not report outcomes according to those used in the full guideline model, some assumption about impact on overall and progression-free survival is required. Since Klotz *et al* showed equivalence of both doses of degarelix compared to leuprolide at 12 months we could (tentatively) assume equivalence in terms of progression-free and overall survival, in which case the cost-effectiveness will be determined by differences in cost alone. The trial does indicate some differences between the three arms in terms of the frequency of adverse events, however these are not statistically significant, and the base case model does not reflect differences between interventions in terms of adverse events in

the palliative treatment section of the model. Thus we assumed that any potential difference in adverse events has no impact on either survival or HRQoL. The drug schedules are the same (a starting dose at time 0, and monthly injections thereafter). Thus the drug cost for the first year of treatment (using BNF prices) will be £3,352, £1,812 and £903 for the degarelix high dose (240mg/160mg thereafter), degarelix low dose (240mg/80mg thereafter) and leuprolide (7.5mg monthly) respectively.

Given the above assumptions, formal modelling is not required to show that a comparison of these three interventions would result in leuprolide being shown to be dominant (i.e. cheaper and equally effective). Therefore, given the limitations in the available evidence, this topic was not evaluated using the full guideline model.

Topic F - Intermittent hormone therapy vs. continuous hormone therapy for men with metastatic prostate cancer

NICE Prostate Cancer Guideline CG58⁵⁵ did not address the question of intermittent versus continuous hormone therapy for patients with metastatic prostate cancer. Two RCTs^{95,122} have shown almost identical survival outcomes, with a slighter longer time to progression in the continuous hormones arm. At the time of analysis only one RCT had published its results in full,⁹⁵ and this data was used in the base case model. No changes to the base case model were required to evaluate this topic.

Topic G - Radium-223 chloride vs. strontium-89 for men with castration-refractory prostate cancer (CRPC) and painful bone metastases

This topic was suggested due to the promising results shown in a Phase III RCT, ALSYMPCA.¹²³ This study suggests that Radium-223 chloride compared to placebo plus best supportive care, including strontium-89, significantly improves overall survival in patients with CRPC that has spread to the bone. Roughly 90% of men with castration refractory disease suffer from painful bone metastases which are currently not treated directly; these patients receive 'best supportive care' which includes strontium-89 to relieve pain.

Strontium-89 is included in the base case model as an additional cost near the end of life, however the health benefit of pain relief is not accounted for in the model. The model structure was therefore adapted to allow for the small survival improvement at the end of the pathway, using the survival difference reported in the ALSYMPCA trial (published as an

abstract only but overall survival curve provided by study PI - personal communication 28/06/12). However, as the Radium-223 chloride does not yet have a list price, this topic was not evaluated using the full guideline model. The structure of the model would easily allow such an evaluation in the future.

Topic H - IMRT and IGRT vs. conformal RT

Intensity modulated radiotherapy and image-guided radiotherapy have been recommended by the Department of Health National Radiotherapy Advisory group (NRAG). Neither intervention was evaluated by NICE in CG58.⁵⁵

A recent HTA report conducted a thorough review of the clinical evidence and found 8 studies reported across 13 publications.¹⁰² The authors concluded that “*the studies are too heterogeneous both for meta-analysis and to attempt to identify variation in effects by dose.*” Given the limitations described, the authors restricted the analysis to those studies that reported biochemical relapse-free survival. They used the outcomes from each study as a different scenario in evaluate in their economic model (see Table 10 below).

Table 10. Modelling scenarios in Hummel et al. 2010¹⁰²

Study	Group	Dose	Scenario
Vora et al. 2007	IMRT	75.6Gy	PSA survival difference
	3DCRT	68 Gy	
Kupelian et al.2002/2005	IMRT	70 Gy*	PSA survival difference
	3DCRT	78 Gy*	
Morgan et al. 2007	IMRT/3DCRT	80/81 Gy	No PSA survival difference

**Considered biologically equivalent*

The third modelling scenario (based on data from Morgan et al 2007) from Hummel et al 2010 was replicated using biochemical recurrence data, frequency of sexual function and urinary adverse events from Widmark et al.⁹⁴ We also included an increased frequency of bowel adverse events with IMRT, as reported in Vora et al.¹²⁴

Topic I - Active surveillance in previously unscreened ‘low risk’ men

This topic indicates that men with low risk disease according to the D’Amico classification are a heterogeneous group and active surveillance may not be the optimal treatment strategy for some of the patients in this group. However the question is vague as it does not propose an alternative risk classification system or an alternative treatment pathway for

men with low-risk disease. Whilst both of these options could in principle be evaluated using the pathway model, no modelling was attempted for this topic as further work would be needed to define an answerable clinical and economic question.

Table 11 summarises the modelling undertaken for each update topic, the structural impact to the base case model and the data requirements for each scenario. Full details of the data used are given in Appendix 5.

Table 11. Summary of amendments to evaluate update topics

Topic	Description	Section of model	Options evaluated in model	Additional evidence required
A	Pelvic radiotherapy with adjuvant hormonal therapy for men with high risk or locally advanced prostate cancer.	Radical treatment >> all patients with 'high-risk' disease	Base case - 50% patients receive HT+RT, 50% patients receive HT alone All patients receive radiotherapy plus hormone therapy All HT alone	none
B	Effective techniques for performing radical prostatectomy.	Radical treatment >> all patients who had surgery in base case model	1. Base case - all patients suitable for surgery receive RRP 2. All patients suitable for surgery receive PRP 3. All patients suitable for surgery receive LRP 4. All patients suitable for surgery receive RALRP	Biochemical recurrence-free survival curve from Martis 2007 (assume equivalence for all 4 surgical interventions). Frequency of three main adverse events RRP, PRP, LRP, RALRP Duration of three main adverse events for each surgical intervention Cost of each surgical intervention
C&D	High dose rate (HDR) brachytherapy in addition to external beam radiotherapy for men with localised or locally advanced prostate cancer. <hr/> Low dose rate (LDR) brachytherapy in addition to external beam radiotherapy for men with localised or locally advanced prostate cancer.	Radical treatment >> all men with intermediate or high risk disease <hr/> Radical treatment >> all men with intermediate or high risk disease	Base case HDR brachytherapy and EBRT LDR brachytherapy and EBRT Brachytherapy only Radiotherapy plus hormone therapy	Biochemical recurrence-free survival curves for HDR + EBRT and LDR + EBRT Frequency of three main adverse events for HDR + EBRT and LDR + EBRT Duration of three main adverse events for HDR + EBRT and LDR + EBRT Cost for HDR + EBRT and LDR + EBRT

Topic	Description	Section of model	Options evaluated in model	Additional evidence required
E	Degarelix (a LHRH antagonist), for men with advanced hormone dependent prostate cancer (locally advanced or metastatic).	Palliative treatment	Not evaluated using full guideline model (insufficient evidence)	PFS not an outcome in key trial. If we assume equivalence with continuous hormones, degarelix would be dominated.
F	Intermittent hormone therapy versus continuous hormone therapy for men with metastatic prostate cancer.	First-line hormone treatment	<ol style="list-style-type: none"> 1. Base case – patients receive continuous hormones, intermittent hormones, bicalutamide monotherapy or bilateral orchidectomy 2. All patients receive continuous hormones 3. All patients receive intermittent hormones 	none
G	Radium-223 chloride versus strontium-89 for men with hormone refractory prostate cancer and painful bone metastases.	Palliative treatment for men with CRPC	Topic not evaluated using full guideline model (no list price for intervention)	Additional overall survival benefit associated with Radium-223. List price for Radium-223
H	Intensity Modulated Radiation Therapy (IMRT) and Image Guided Radiation Therapy (IGRT) as an alternative to conventional therapy for men undergoing radiation treatment.	Radical treatment	<ol style="list-style-type: none"> 1. Base case all patients receive 3D-conformal radiotherapy (3DRT) 2. All patients receive IMRT <p><i>Note: IGRT was not evaluated</i></p>	<p>Biochemical recurrence-free survival curves for IMRT and 3DRT, from the three sources (Vora, Kupelian & Morgan)</p> <p>Frequency of bowel dysfunction for the two combination treatments (other adverse events assumed same as RT in the base case)</p> <p>Duration of bowel dysfunction for IMRT and 3DRT</p> <p>Cost of the two combination treatments</p>
I	Active surveillance in previously unscreened 'low risk' men.	Classification of risk – covers both diagnosis and imaging & radical treatment	Topic not evaluated using full guideline model (weakly defined question)	none

Results

Base case estimates of costs and health outcomes

The results of the base case model provide a mean estimate of the numbers of patients in each section of the pathway, and the associated mean costs and mean health effects (life years and QALYs gained) for the total cohort of patients in the model. Table 12 shows the estimated number of men in each section of the base case (current service) model, based on a cohort of 200,000 men referred into secondary care with suspected prostate cancer. The analysis suggests that just over 20% of men presenting to secondary care with symptoms suspicious of prostate cancer will be diagnosed with prostate cancer.

On average, around 1 in 3 men diagnosed with prostate cancer will receive 'watchful waiting' and 1 in 10 will receive active surveillance. Approximately 60% of men diagnosed with prostate cancer will receive some form of radical treatment, including those men who switch to treatment after some time of active surveillance. Approximately half of all men diagnosed with diagnosed prostate cancer will at some point in their lives receive hormone treatment. Around one third of men diagnosed with prostate cancer are expected to receive palliative chemotherapy.

Table 12. Intermediate results (per 1,000 men referred)

Intermediate results	2.5 th percentile	97.5 th percentile	Mean
Number patients diagnosed	147	293	216
Number patients never diagnosed	707	853	784
Number patients entering watchful waiting	39	90	62
Number patients entering active surveillance	14	30	21
Number patients undergoing radical treatment	87	175	127
Number patients getting hormone treatment	70	139	103
No. patients receiving 1 line of palliative treatment	9	19	13
No. patients receiving 2 lines of palliative treatment	3	14	7
No. patients receiving 3 lines of palliative treatment	2	8	4
No. patients receiving 4 lines of palliative treatment	53	106	79

Results from probabilistic base case model.

Table 13 shows the contribution of each segment of the model to overall life years gained and QALYs gained. The results show that on average, each man referred into secondary care

with suspected prostate cancer can expect to live for 13.96 years and will accrue around 10.96 QALYs (undiscounted).

Table 13. Health outcomes by base case model segment (per 1,000 men referred)

Model segment	Undiscounted		Discounted	
	LYGs	QALYs	LYGs	QALYs
From model entry to death or diagnosis	11070	9008	7934	6456
From start of radical treatment to cure, relapse or other cause death	1199	934	897	699
Active surveillance to radical treatment, death or watchful waiting	163	133	131	107
Watchful waiting to palliative treatment or death	539	439	411	335
Palliative treatment to the end of PFS3	336	247	263	193
4th line palliative treatment: chemotherapy	55	35	40	25
From end of PFS from 4th line treatment to death	603	340	372	210

Results from probabilistic base case model

The costs associated with each workcentre within the model are shown in Table 14. The model suggests that over the lifetimes of a cohort of 1,000 men, expenditure on radical treatment is almost three times that of palliative treatment. For the 1,000 men referred, the expected discounted lifetime cost associated with the base case configuration of UK services is around £6.5 million, £6,500 for each man referred .

Table 14. Costs from the base case model (per 1,000 men referred)

Workcentre	Undiscounted costs	Discounted costs
PSA secondary care	£2,342,409	£1,682,742
GP PSA test	£374,859	£268,533
Biopsy	£390,984	£354,935
Bone scan	£1,158	£967
BPH	£272,723	£247,330
Determine treatment	£74,719	£73,995
Watchful waiting	£25,629	£19,291
Active surveillance	£52,332	£42,634
Radical treatment	£3,731,717	£2,880,556
Palliative treatment	£1,097,857	£933,514
Total	£8,364,387	£6,504,496

Results from probabilistic base case model

Incremental cost-effectiveness analysis

Table 15 presents the mean service costs and health outcomes for all thirteen possible variations to the pathway, based on 1,000 patients. A full incremental analysis was undertaken within each guideline topic; these results are described below.

Table 15. Summary of mean costs and outcomes for all options tested (per 1,000 men referred)

Options	Description	Undiscounted Mean LYGs	Undiscounted Mean QALYs	Undiscounted Mean Cost	Discounted Mean LYGs	Discounted Mean QALYs	Discounted Mean Cost
Base case	Current recommended service pathway	£13,965	£10,963	£8,364,387	£10,049	£7,902	£6,504,496
A2	All patients with high risk disease receive RT plus HT	£14,038	£11,036	£8,517,120	£10,084	£7,941	£6,554,880
A3	All patients with high risk disease receive HT only	£13,894	£10,892	£8,212,932	£10,014	£7,862	£6,454,294
B2	All patients suitable for surgery have PRP	£13,965	£10,962	£8,390,938	£10,049	£7,901	£6,522,135
B3	All patients suitable for surgery have LRP	£13,965	£10,962	£8,408,130	£10,049	£7,901	£6,537,235
B4	All patients suitable for surgery have RALRP	£13,965	£10,969	£8,296,090	£10,049	£7,906	£6,462,302
C/D2	All patients not suitable for surgery have HDR-BT plus RT	£13,948	£10,966	£9,101,662	£10,039	£7,903	£7,214,983
C/D3	All patients not suitable for surgery have LDR-BT plus RT	£14,014	£11,037	£8,813,412	£10,072	£7,946	£6,865,768
C/D4	All patients not suitable for surgery have BT	£14,075	£11,104	£7,380,321	£10,099	£7,977	£5,716,053
C/D5	All patients not suitable for surgery have RT plus HT	£14,028	£11,020	£8,778,727	£10,080	£7,933	£6,741,876
F2	All patients receive continuous HT as 1 st line palliative treatment	£14,023	£10,994	£8,445,704	£10,084	£7,921	£6,560,991
F3	All patients receive intermittent HT as 1 st line palliative treatment	£13,895	£10,921	£8,391,740	£10,017	£7,880	£6,518,776
H2	All patients receive IMRT instead of conformal RT plus HT	£13,965	£10,956	£8,521,166	£10,049	£7,897	£6,614,686

Topic A - Pelvic radiotherapy with adjuvant hormone therapy for men with localised prostate cancer

Three possible alternatives were compared in Topic A. In the base case model we assumed that 50% of men with high risk or locally advanced disease would receive radiotherapy with adjuvant hormone treatment, whilst the remaining 50% would receive hormone treatment alone. Option A2 assumed that all eligible men would receive radiotherapy with adjuvant hormone treatment. Option A3 assumed that all eligible men would receive hormone treatment alone.

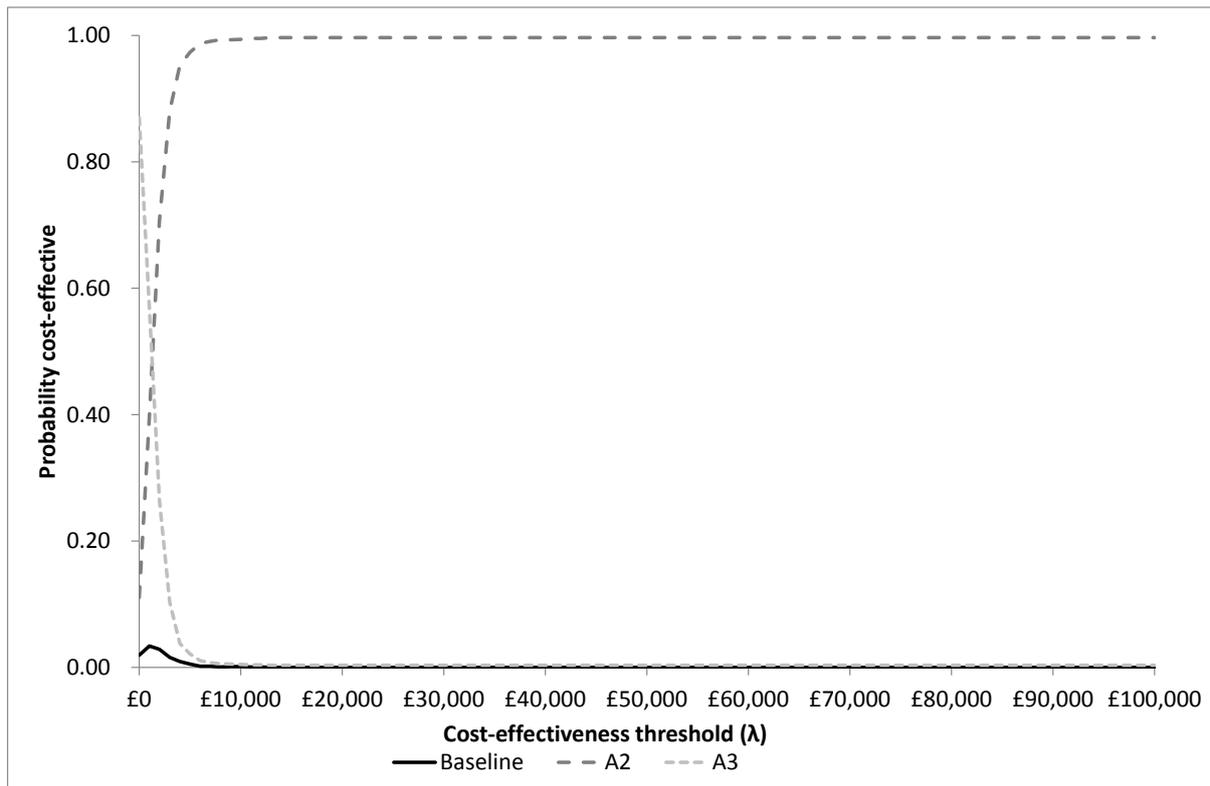
The results of the analysis are shown in Table 16. This suggests that offering all men with high risk or locally advanced disease radiotherapy in addition to hormone treatment, rather than hormone treatment alone is expected to be the most effective and the most expensive option. Offering hormone therapy alone is expected to produce the fewest QALYs and the lowest overall cost. The base case service, which involves a combination of the other two options, is ruled out due to extended dominance. Radiotherapy plus hormone therapy versus hormone therapy alone is expected to yield a discounted incremental cost-effectiveness ratio of around £1,275 per QALY gained.

Table 16. Incremental cost-effectiveness results - Topic A (per 1,000 men referred)

Option	Total QALYs	Total cost	Inc. QALYs	Inc. cost	ICER
RT plus HT (A2)					1269
	7941	6554880	39.72	50384	
Base case (combination)	7902	6504496	39.16	50202	ext dom
HT only (A3)	7941	6554880	-	-	-

Figure 10 presents cost-effectiveness acceptability curves for the three options compared in Topic A. At very low values of λ (when health is valued less), hormone therapy alone is expected to produce the greatest net benefit. At threshold values of around £5,000 or more, radiotherapy plus hormone therapy is expected to produce the most net benefit. Assuming a willingness-to-pay threshold of £20,000 per QALY gained, the probability that radiotherapy plus hormone therapy produces the greatest net benefit is approximately 1.0.

Figure 10. Cost-effectiveness acceptability curves - Topic A



Topic B - Surgical techniques for localised prostate cancer: open radical retropubic prostatectomy (RRP), trans-perineal prostatectomy (PRP), laparoscopic prostatectomy (LRP) or robot assisted laparoscopic prostatectomy (RALRP)

Topic B involves a comparison of four alternative surgical procedures for men eligible to undergo radical prostatectomy. The base case strategy assumed all men would undergo a standard open procedure. Option B2 assumed that all men would have a transperineal radical prostatectomy (PRP). Option B3 assumed men would have a laparoscopic radical prostatectomy (LRP). Option B4 assumed that all men would receive robotic-assisted laparoscopic surgery (RALRP).

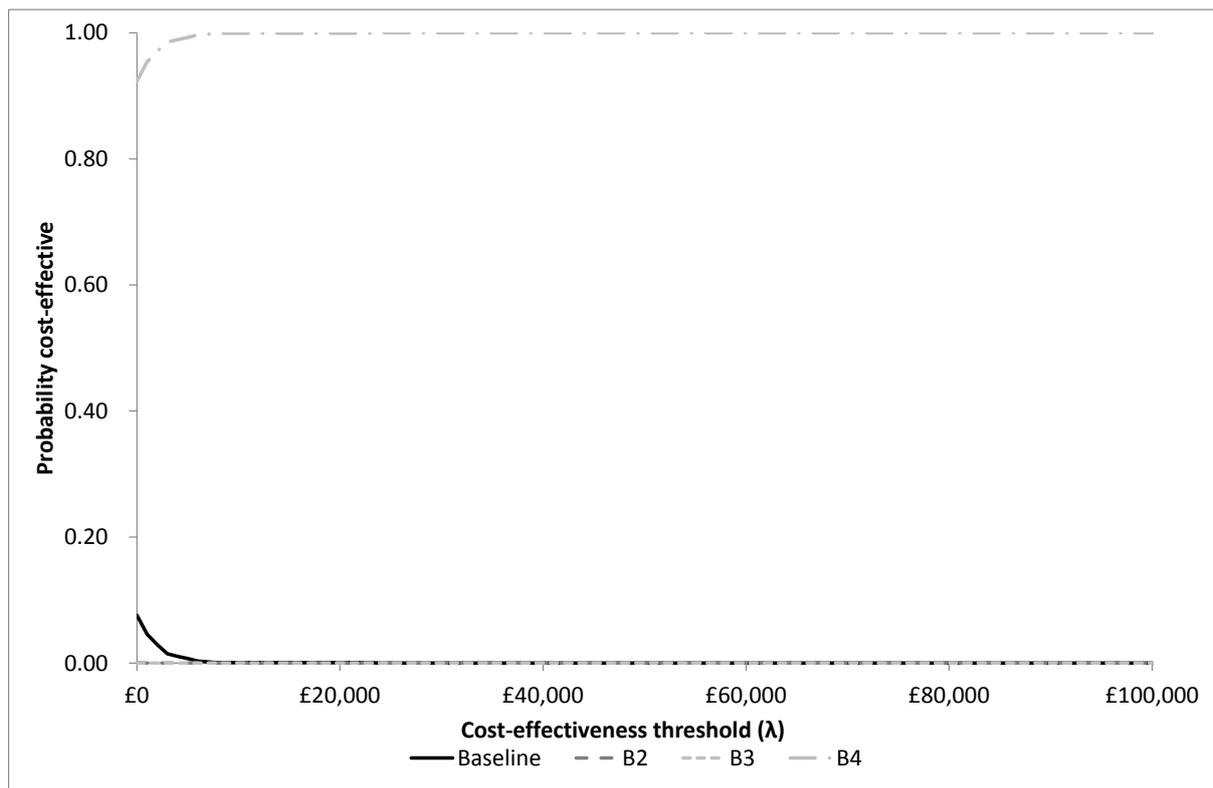
The results of the economic analysis of this topic are presented in Table 17. Unsurprisingly, the model results indicate very little difference in terms of incremental health gains between the evaluated options. The model suggests that RALRP (Option B4) is associated with a slight increase in QALYs compared to the other options. This is also the least expensive option, hence it is expected to dominate all other options.

Table 17. Incremental cost-effectiveness results - Topic B (per 1,000 men referred)

Option	Total QALYs	Total cost	Inc. QALYs	Inc. cost	ICER
RALRP (B4)	7906	6485560	4	-18936	Dominating
Base case (option)	7902	6504496			Dominated
PRP (B2)	7901	6522135			Dominated
LRP (B3)	7901	6537235			Dominated

Figure 11 presents cost-effectiveness acceptability curves for the four options compared in Topic B. Assuming a willingness-to-pay threshold of £20,000 per QALY gained, the analysis shows that RALRP is always likely to produce the greatest net benefit, compared to the standard RRP open procedure. However there is still considerable structural uncertainty with respect to the duration of adverse events and the costs of managing these, which are not addressed with probabilistic sensitivity analysis.

Figure 11. Cost-effectiveness acceptability curves - Topic B



Topic C/D - Brachytherapy +external-beam radiotherapy for men with localised or locally advanced prostate cancer

The evaluation of Topic C/D involved a comparison of five alternative options. The base case model assumes that patients with intermediate risk disease will receive radiotherapy plus hormone treatment and those with high risk disease may receive radiotherapy plus hormones or brachytherapy monotherapy. Option C/D2 involves brachytherapy in combination with external beam radiation as high-dose. Option C/D3 involves brachytherapy in combination with external beam radiation as low-dose. Option C/D4 represents brachytherapy monotherapy. Option C/D5 represents radiotherapy plus adjuvant hormone treatment.

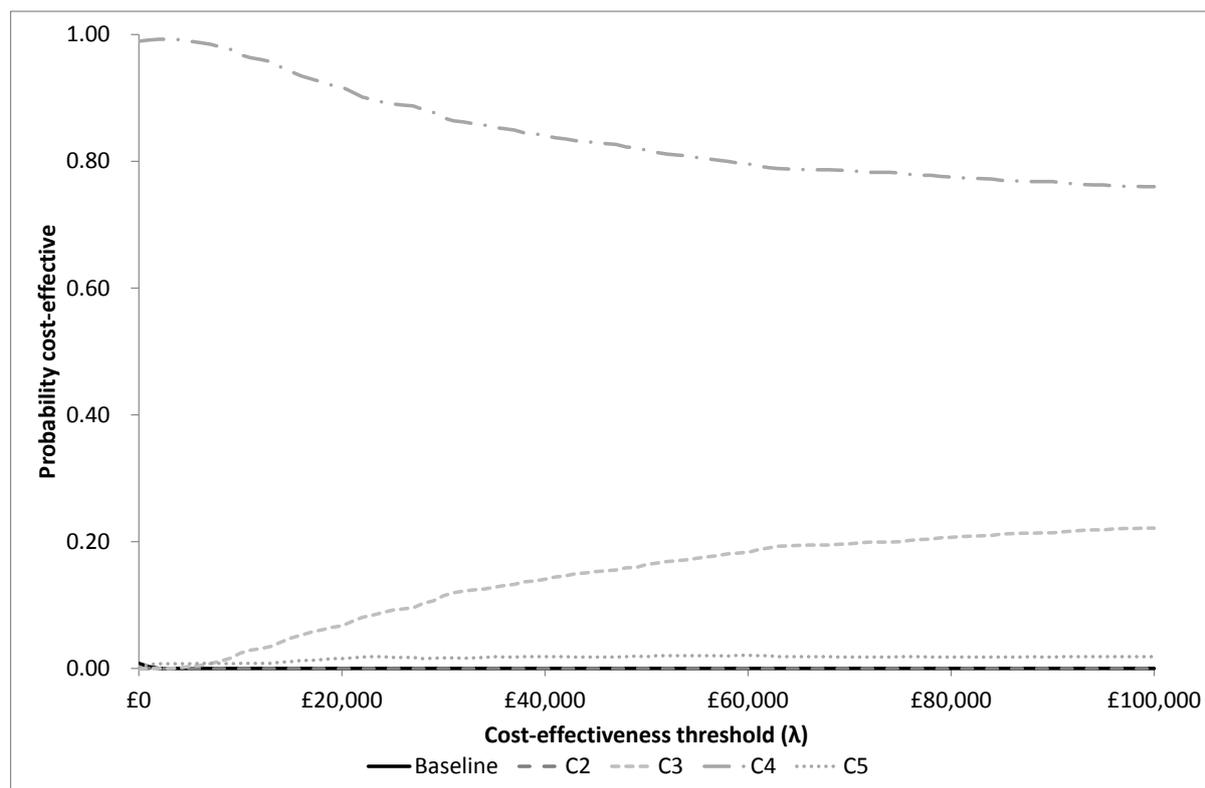
Table 18 presents the results for the economic analysis of Topic C/D. The results suggest that brachytherapy monotherapy (Option C/D4) is associated with the highest expected QALY gain and the lowest cost. All other options, including the current base case, are dominated by this strategy.

Table 18. Incremental cost-effectiveness results - Topic C (per 1,000 men referred)

Option	Total QALYs	Total cost	Inc. QALYs	Inc. costs	ICER
Brachytherapy mono (C/D4)	7977	5,716,053			Dominating
Brachytherapy + LD EBR (C/D3)	7946	6,865,768	-	-	Dominated
RT + adjuvant HT(C/D5)	7933	6,741,876	-	-	Dominated
Brachytherapy + HD EBR (C/D2)	7903	7,214,983	-	-	Dominated
Base case	7902	6,504,496	-	-	Dominated

Figure 12 presents the cost-effectiveness acceptability curves for Topic C/D. Assuming a willingness-to-pay threshold of £20,000 per QALY gained, the probability that brachytherapy monotherapy produces the greatest net benefit is approximately 0.84.

Figure 12. Cost-effectiveness acceptability curve - Topic C



Topic F - Intermittent hormone therapy vs. Continuous hormone therapy for men with metastatic prostate cancer

Topic F involved the economic comparison of three options. The base case model assumed 90% of men who received either continuous or intermittent hormone treatment would receive LHRHa continuously, with 10% receiving LHRHa intermittently. Option F2 represents continuous hormone therapy and Option F3 represents intermittent hormone therapy.

The results of the economic analysis of Topic F are presented in Table 19. These suggest that continuous hormone treatment is expected to produce the greatest number of QALYs at the highest cost. Option F3 (intermittent hormones) is expected to be dominated by the base case service. Continuous HT is expected to cost approximately £3,000 per QALY gained when compared against the base case service.

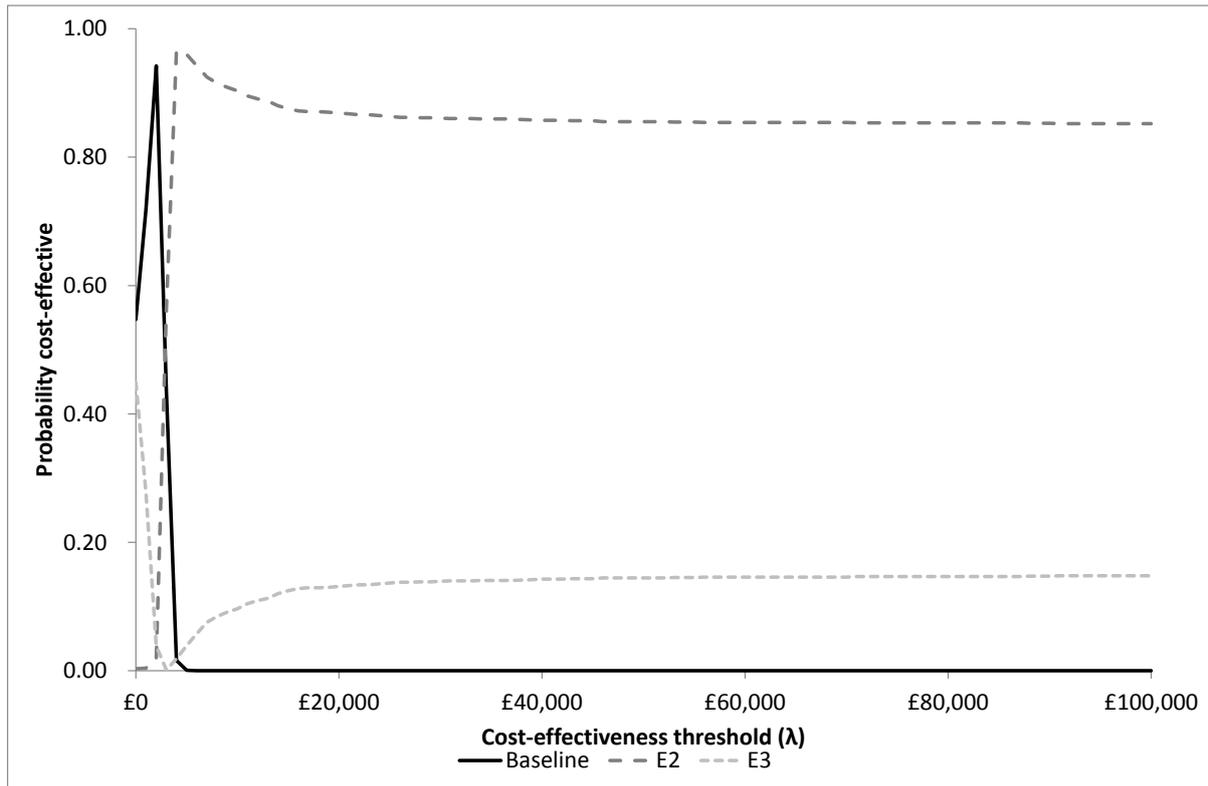
Table 19. Incremental cost-effectiveness results - Topic F (per 1,000 men referred)

Option	Total QALYs	Total cost	Inc. QALYs	Inc. costs	ICER
Continuous HT (F2)	7921	£6,560,991	19	56495	£2,963
Base case	7902	£6,504,496	-	-	-

Intermittent HT (F3)	7880	£6,518,776	-	-	Dominated
-----------------------------	------	------------	---	---	-----------

Figure 13 presents the cost-effectiveness acceptability curves for Topic F. Assuming a willingness-to-pay threshold of £20,000 per QALY gained, the probability that continuous hormone therapy produces the greatest expected net benefit is approximately 0.86.

Figure 13. Cost-effectiveness acceptability curve for Topic F



Topic H - IMRT and IGRT vs. conformal RT

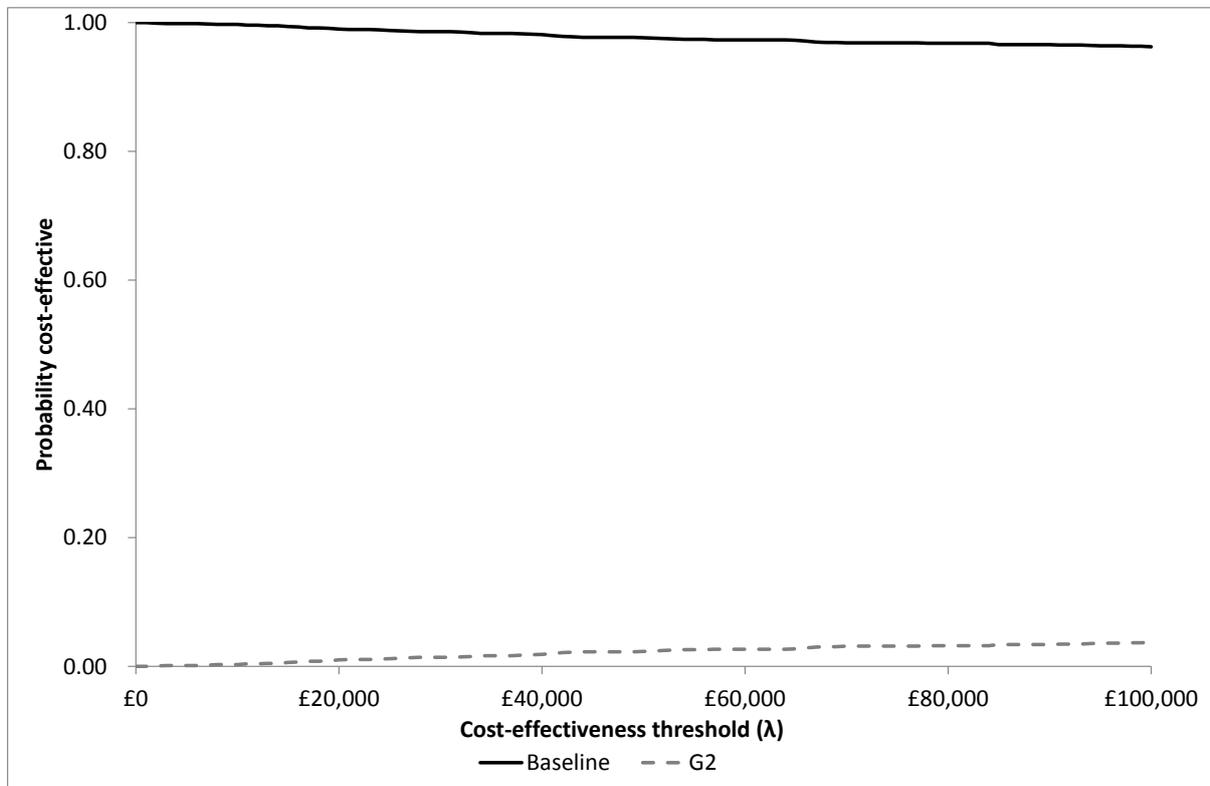
Topic I compares intensity-modulated radiotherapy (Option H2) against the base case strategy - 3D-conformal radiotherapy. The headline economic results are presented in Table 20. The economic analysis suggests that IMRT is expected to result in fewer QALYs and a greater expected cost than 3D conformal radiotherapy.

Table 20. Incremental cost-effectiveness results - Topic H (per 1,000 men referred)

Option	Total QALYs	Total cost	Inc. QALYs	Inc. costs	ICER
Base case (3D conformal RT)	7902	£6,504,496			Dominating
IMRT (H2)	7897	£6,614,686	-	-	-

Figure 14 presents cost-effectiveness acceptability curves for Topic H. Assuming a willingness-to-pay threshold of £20,000 per QALY gained, the probability that 3D conformal radiotherapy produces more net benefit than IMRT is approximately 0.98.

Figure 14. Cost-effectiveness acceptability curves - Topic H



Summary of economic results and implications for updating guideline topics

Table 21 summarises the results of the economic analyses in terms of the expected absolute net benefit for each option and the probability that each option produces the greatest net benefit as compared against other options within each guideline topic. The option which is preferred on the grounds of expected cost-effectiveness within each topic is highlighted in bold.

For Topics A, B, C/D and F, the economic analysis indicates that the current base case service is *not* expected to produce the greatest amount of net benefit. In each of these circumstances, additional health benefits may be attainable by pursuing other treatment options. This would indicate that these topics should be considered for update in the guideline. On the basis of the magnitude of expected incremental net benefit lost between each option as compared against the base case service, Topics A, C/D and F represent the

top three priorities for update on economic grounds. Topic B is also associated with lost net benefit, although this is less than that for the other topics.

Table 21. Summary of net benefit results (per 1,000 men referred)

Topic	Option	At £20,000 per QALY gained		At £30,000 per QALY gained	
		Incremental Net benefit [§]	Probability optimal	Incremental net benefit [§]	Probability optimal
A	Base case	£0	0.00	£0	0.00
	A2 - RT plus HT	£743,915	1.00	£1,141,064	1.00
	A3 - HT only	-£733,017	0.00	-£1,124,627	0.00
B	Base case	£0	0.00	£0	0.00
	B2 – PRP	-£32,030	0.00	-£39,225	0.00
	B3 – LRP	-£50,768	0.00	-£59,783	0.00
	B4 - RALRP	£102,385	1.00	£144,110	1.00
C/D	Base case	£0	0.00	£0	0.00
	C2 - Brachytherapy + HD EBR	-£679,666	0.00	-£664,255	0.00
	C3 - Brachytherapy + LD EBR	£530,558	0.14	£976,473	0.16
	C4 - Brachytherapy mono	£2,305,934	0.84	£3,064,680	0.82
	C5 - RT + adjuvant HT	£391,500	0.02	£705,940	0.02
	F	Base case	£0	0.00	£0
E2 - Continuous HT		£324,901	0.86	£515,599	0.86
E3 - Intermittent HT		-£443,495	0.14	-£658,103	0.14
G	Base case	£0	0.98	£0	0.98
	G2 - IMRT	-£200,262	0.02	-£245,297	0.02

[§]The potential net benefit gain from changing the pathway compared to base line model, per 1,000 men referred

Discussion

The full guideline model presented within this chapter captures the key events, costs and health outcomes associated with the main elements of care for men referred into secondary care with suspected prostate cancer. The model reflects the broad range of components of the care pathway including diagnosis and imaging, GP monitoring, treatment planning, watchful waiting, active surveillance, radical treatments, follow-up care and palliative

treatments. The full guideline model differs from conventional piecewise models in that it adopts a broader pathway-level scope whilst retaining a high level of depth across the individual pathway components. Whilst most conventional models are developed to address a single decision problem at a specific decision point in a care pathway, this full guideline model provides a platform for the evaluation of multiple options for service change across the whole service pathway. Whilst these are presented solely as analyses of individual guideline topics, the model also has the functionality to evaluate multiple topics simultaneously. This represents a more powerful decision-making tool than has been used in the majority of existing clinical guidelines.

Headline probabilistic model results

We evaluated six of the nine selected guideline topics. Our analysis indicates that for five of these topics, the current guideline recommendations are *not* expected to produce the greatest net benefit. Although these results are not definitive – as, for example, they are not based on systematic reviews of the evidence - they are indicative of areas where further investigation is likely to be of value. In particular, the economic analysis indicates that:

- offering all men with high risk or locally advanced disease radiotherapy plus hormone treatment is expected to have an incremental cost-effectiveness ratio of around £1,275 per QALY gained when compared against the hormone therapy alone;
- Robot-assisted laparoscopic prostatectomy (RALRP) is expected to dominate all other surgical options for localised prostate cancer;
- Brachytherapy monotherapy is expected to dominate alternative radiotherapy options men with localised or locally advanced prostate cancer, and;
- Continuous hormone treatment is expected to yield an incremental cost-effectiveness ratio of around £2,963 compared against the current mix of continuous and intermittent hormone treatments.
- 3D conformal radiotherapy is expected to dominate intensity-modulated radiotherapy.

In terms of net benefit lost (by choosing the base case service over other potentially more cost-effective decision alternatives), the following three topics represent the highest priorities for update:

- Topic A - Pelvic radiotherapy with adjuvant hormone therapy for men with localised prostate cancer
- Topic C/D - Brachytherapy plus external-beam radiotherapy for men with localised or locally advanced prostate cancer
- Topic F - Intermittent hormone therapy versus continuous hormone therapy for men with metastatic prostate cancer.

Evidently, there is disagreement between the topics which would be prioritised on economic grounds and those which would be prioritised by the stakeholders who responded to our survey (see Chapter 3). Whilst both the economic analysis and the surveys indicate that some benefit may be obtained by prioritising Topic B (surgical techniques), this is associated with a comparatively smaller amount of net benefit lost relative to the base case service.

Limitations of the analysis

As with any health economic model, the credibility of this model and its results are largely dependent on the quality of the evidence used to inform it. There are a number of limitations of the economic analyses presented here, the majority of which derive from limitations in the available evidence base. It is important to recognise that most of these problems are not a result of the modelling methodology itself; rather, the same problems with evidence would apply to the development of any health economic model, irrespective of its scope. One of the key values of mathematical modelling, in particular modelling on this scale, is its ability to draw out the key gaps and uncertainties in the available evidence base. The following key simplifications should be borne in mind when interpreting the results of the economic analysis:

- The available registry data did not include the PSA score at diagnosis. Gleason score is assumed to be fixed from the point of diagnosis when in reality this could change following a repeat biopsy.
- PSA-based criteria for informing treatment decisions were not fully captured in the active treatment portions of the model.
- The potential impact of misclassification of diagnostic tests was not reflected in the model because of the inherent difficulties of modelling inaccurate diagnoses and the

impact on outcomes. In addition, test operating characteristics were captured only for TRUS.

- The model includes only a partial representation of disease natural history – the model do not include the incidence of prostate cancer over time -that is, all patients either have or do not have prostate cancer upon model entry.
- Much of the evidence used to inform the treatment portions of the model required naïve indirect comparisons due to a lack of randomised evidence.
- We assumed that biochemical progression and disease recurrence have an equivalent impact upon clinical decision-making and subsequent prognosis.
- Survival benefits for sequences of palliative treatments were assumed to be driven by the first-line treatment in the sequence. In addition, sequences of palliative treatments are modelled according to an overall mean time and do not fully reflect treatment variations between patients.

Key evidence limitations and model simplifications

The adoption of an individual patient-level simulation approach can place increasingly heavy demands on the model in terms of data. In the absence of well reported summary statistics, which include variance-covariance matrices across multiple patient characteristics, access to individual-level data on patient characteristics at model entry is essential to fully characterise the correlations between the key patient characteristics. We used UK cancer registry data on age, clinical stage and Gleason score from the SWPHO registry. This registry dataset did not however include information on patients' PSA scores; instead these were "back-calculated" conditional on the characteristics for which we had data. Furthermore, we did not identify any robust evidence concerning the relationship between PSA score, underlying disease progression and treatment. We also necessarily assumed that Gleason score was fixed from the point of diagnosis (we found no evidence to reflect the potential trajectory of change in Gleason score over time). Given these issues, the lack of evidence invariably limits the level of depth (or detail) reflected in the pre-diagnostic portion of the model. The consequence of this lack of evidence led to certain elements of the model becoming "blunt" and in some instances leads to a separation between our conceptual understanding of how diagnostic and treatment decisions are made in practice and the extent to which the model can reflect these decisions. As a consequence, we were unable to

capture any of the NICE prostate cancer guideline CG58 recommendations based on observed changes in PSA score, doubling time or velocity.

In addition, the simulation model includes only a partial representation of the natural history of prostate cancer. As a result, the portion of the model dealing with the underlying natural history and diagnosis is fairly simplistic. This set of simplifications was driven by significant limitations in the available evidence base with respect to underlying natural history progression and the lack of good quality evidence relating to the probabilities and consequences of incorrect diagnostic decision-making. We did not assess the impact of the error associated with PSA testing, MRI imaging or bone scans. We also assumed TRUS biopsy was associated with perfect specificity and the evidence used to provide evidence of the false negative rate is dated.⁸⁹ As with the evaluation of any diagnostic intervention, the lack of evidence regarding the costs and consequences of counterfactual pathways that would be followed by patients with misclassified disease presents a further challenge which we chose not to fully address within the model.

Our estimation of the natural history of prostate cancer was crude, calibrated using data on patients in the watchful waiting arm of the Bill-Axelson *et al* RCT⁷⁷ and UK-specific life tables using the Metropolis-Hastings algorithm. Whilst we believe the calibration method is appropriate, undoubtedly other information from further sources tells us something about these unobservable parameters. This could include evidence from screening trials, autopsy studies or evidence from prostate cancer surveillance and monitoring studies. The design and implementation of a more comprehensive calibration process would have increased the robustness of the natural history estimation, but would have required considerable additional effort and resource. Given that none of the topics selected for evaluation actually related to diagnostic interventions or screening, this additional effort could be argued to have at best a limited payoff for the context of the case study. However, it is acknowledged that the value of explicitly modelling epidemiology and natural history progression may be greater in guidelines for other diseases. Further extension of this component of the model may increase the utility of the model in addressing other decision problems elsewhere in the prostate cancer pathway.

The treatment portion of the model is also subject to a number of problems relating to the availability and quality of evidence. No modelling approach can reconcile the absence of

head-to-head trials comparing all relevant radical treatment options and a comprehensive evidence network. In most instances, we had little choice but to use naïve indirect comparisons to capture the relative effects of radical and palliative treatments. This breaks randomisation between studies and can lead to significant bias and confounding. However, again, we believe this problem lies in the evidence base rather than the modelling approach *per se*. In other instances, we were also limited by relevant trials reporting less relevant or useful outcomes. The palliative treatment portion of the model is intended to reflect the impact of different sequences of treatments on HRQoL and survival. However, we did not identify any studies which assessed planned sequences of treatment. As such we assumed that the first-line treatment determines the survival benefit for the sequence, with subsequent treatments influencing the amount of time for which the patient is progression-free. This is a common problem in cancer evaluations and is again not specific to this particular modelling methodology. In addition, because progression-free survival includes survival as an event, we were unable to reflect first-order uncertainty in this part of the model.

Usefulness of the broader modelling approach

Volume of economic evidence generated

A total of nine topics were selected for evaluation. Two of these topics dealt with potentially competing interventions at the same point in the pathway (Topics C and D). Six topics were subjected to formal cost-effectiveness analyses using the full guideline model. The economic analysis of three topics was not attempted. Topic E was not evaluated due to a total lack of evidence, Topic G because the intervention did not have a list price (and is not being actively marketed), and Topic I was not undertaken as the question was not sufficiently defined to identify an intervention.

It is reasonable to suggest that the full guideline model provides considerably more economic information than would otherwise be available from the conventional piecewise approach. It remains unclear however, whether the resources and effort required to develop the full guideline model exceeds what would be required to undertake the same economic analyses using five individual *de novo* piecewise models.

With the exception of Topic I (active surveillance strategies for previously unscreened men), all of the topics selected for the case study reflect active treatments for diagnosed disease. In principle, the full guideline model could also be used to address a wide range of other decision problems across the prostate cancer service which were not selected for inclusion in the update (for example, assessing the optimal frequency of GP monitoring visits or assessing alternative biopsy techniques).Development time

The time and resource required to develop the model were considerable. Model development began in August 2010 and the final results were produced in November 2012. Two modellers were involved in designing and implementing the model. It is reasonable to suggest that a considerable proportion of this time involved developing familiarity with the software package and the inevitable learning curve associated with developing models on this scale. It may have been possible to develop the same model within the timescales of a 'live' clinical guideline although this could represent a risk to the delivery of the guideline. The magnitude of this risk will inevitably vary across different guidelines and different disease areas. Alternatively, it may be possible to develop this type of model before the clinical guideline development process begins.

Problems of the approach

Whilst the adoption of a broad model scope is attractive in terms of the volume and consistency of economic evidence that can be generated using a single model, it does carry with it a number of potential risks and costs. For example, one could argue that the scope was too broad – we modelled the breadth of the whole pathway from secondary care referral yet only topics related to treatment were evaluated using the model. Thus, considerable development time was devoted to developing parts of the model for which topics were not actually prioritised. Of course, the case study was not undertaken a part of a live guideline process and the model may have potential for evaluating a wider range of topics than those selected for inclusion in the case study.

The development of a single model may offer consistency but also carries a cost in terms of running the evaluation. It remains debateable as to whether simulation models are easier or harder to check than cohort-based models. However, when an error is identified, either in the conceptual or quantitative basis of the model, this error will influence all decisions problems addressed using that model. Where errors are spotted, this can mean re-running

all analyses, potentially multiple times. Where errors are not spotted, these may permeate through the evaluation of multiple topics (although the precise nature of the error will determine whether this makes a difference to the conclusions of any individual topic). Given the likely computational burden associated with this type of model, this can represent a negative aspect of the approach. In this case study, all analyses were re-run five times each of which required approximately 1,200 computation hours. Whilst this was not ideal, it was possible by spreading the model runs across multiple computers using multi-threading; whilst not substantial in this case, pursuing this type of modelling approach may have implications for purchasing both hardware and software.

A final potential problem relates to how this type of large-scale complex model would be interpreted by a GDG. We did not have access to the GDG itself and so we were unable to gauge whether they would find this type of model more or less useful than conventional piecewise models. Testing of the approach, and the qualitative elicitation of the views of GDG members should be considered a priority for future research.

Chapter 5. Case study 2: Full guideline model for atrial fibrillation

Introduction

This chapter describes the development of a Discrete Event Simulation (DES) intended to reflect the course of atrial fibrillation for a representative UK cohort of patients diagnosed and treated in accordance with the service pathway recommended in NICE Clinical Guideline CG36.¹²⁵ The model was designed to predict the incidence of AF-related risks and associated health outcomes and expenditure, and to estimate the cost-effectiveness of some possible changes to the currently-recommended pathway.

The Background section below provides a brief introduction to AF and the NICE clinical guideline. The Methods section then describes how the model was developed, tested and used. It starts with an explanation of how a preliminary literature review provided an initial understanding of the condition, its treatment and the nature of evidence likely to be available to support a model. The process that was used to develop the conceptual design of the model is then outlined, and the key elements of the boundary and scope of the model, the service pathway and the disease process are described. This is followed by a description of how the conceptual model was implemented as a DES in SIMUL8 software, and the sources of data used to parameterise the model and the methods of verification and validation are described. The final part of the Methods section presents the approaches used to investigate the potential update topics defined in Chapter 3. Estimates of the cost-effectiveness of a series of possible changes to the service pathway related to these topics are presented in the Results section. This starts within comparisons of 'mutually exclusive' options within each topic, and then considers possible interactions between options across the topics. In the Discussion, we reflect on the experience of this case study, and consider what it tells us about the motivating questions for the project: the feasibility and potential usefulness of full guideline models as a means of conducting economic evaluation in NICE guidelines.

Background

Atrial fibrillation (AF) is a condition characterised by irregular and rapid heart rhythm.^{126;127;127} It can cause a range of symptoms including chest pains, palpitations,

angina, shortness of breath and fatigue, and can sometimes present as a critical condition with haemodynamic instability requiring urgent treatment, although in contrast some patients do not experience symptoms at all and might be unaware of their condition. AF is associated with a greatly increased risk of death from stroke and other thromboembolic events, heart failure, and cardiovascular disease. Three types of AF have been distinguished; paroxysmal, persistent and permanent AF. Paroxysmal AF is characterised by short episodes of irregular heart rhythm lasting less than 7 days, normally less than 48 hours. Persistent AF is associated with longer episodes, which do not terminate without intervention. In permanent AF there is a perpetual fibrillation of the atria. The 'natural' course of AF is generally progressive, with the frequency and duration of symptomatic episodes usually increasing over time.

AF is common, affecting 1-2% of the general population, and is associated with age; a European study estimated the prevalence at 0.7% for people aged 55-59, rising to 17.8% for people aged over 85.¹²⁷ Recent increases in AF prevalence have been attributed to improvements in survival for cardiovascular conditions associated with AF, and the ageing population. Resources consumed in the treatment of AF are estimated to account for nearly 1% of the UK NHS expenditure. The impact of AF on mortality and quality of life, and the associated economic burden led to the commissioning of a Clinical Guideline by NICE.

The scope for the NICE Clinical Guideline (CG36), published in 2006, covered the processes of patient care from identification and diagnosis, treatment for the prevention of stroke and thromboembolism, electrical and pharmacological methods to correct heart rhythm ('cardioversion' to achieve sinus rhythm), drugs to maintain heart rhythm or to control heart rate, monitoring and referral for specialist electrophysiological interventions such as pacing or ablation.¹²⁸ The guideline also covered acute treatment for haemodynamically unstable patients, and the prevention and treatment of AF in patients undergoing cardiac surgery.

Methods

Preliminary literature review

To inform model development, we conducted an initial review of literature on published economic models for the disease area, related models from NICE guidance (e.g. technology appraisals) and other HTA bodies and guideline developers. We searched the following

secondary databases, using general disease/patient group search terms: (a) CRD NHS Economic Evaluation Database; (b) CRD HTA Database; (c) NHS Evidence; and, (d) Guidelines International Network database. This search was intended as a rapid means of identifying appropriate model structures and sources of data. We did not conduct formal critical appraisal of published economic evaluations, or summarise their findings.

Several documents were identified that were very influential in the development of our model structure. These included economic evaluations and models that covered different aspects of the diagnosis and treatment of AF that we sought to bring together in a model of the whole service pathway and disease process:

- **Case finding:** An HTA funded project by Hobbs *et al*¹²⁹ included a clinical trial and economic evaluation of methods for screening for AF. This provided data on the accuracy of diagnostic methods, as well as informing the design of the decision tree in the diagnostic section of our model.
- **Antithrombotic therapy:** Various models have been developed to evaluate the cost-effectiveness of antithrombotic therapy.¹³⁰⁻¹³⁵ Drug treatments include antiplatelet agents (aspirin and clopidogrel) and anticoagulants (warfarin, dabigatran and rivaroxaban). These drugs are all effective at reducing the risk of thromboembolism, but at the risk of causing dangerous bleeds. In addition, the mainstay of oral anticoagulation (OAC), warfarin, requires regular monitoring which is difficult, inconvenient for the patient and expensive. The available models estimate the balance between these various risks and their health and financial consequences. For this element of our model, we drew particularly on the models developed for the recent NICE Technology Appraisals of dabigatran and rivaroxaban, and the critique of these models provided by the Evidence Review Groups and the Committee considerations.^{130;131;136-139}
- **Antiarrhythmic therapy:** Another recent NICE Technology Appraisal that provided valuable information for the construction of our model was TA197, which compared dronedarone with other drugs for the maintenance of sinus rhythm (amiodarone and the Class 1c antiarrhythmic drugs (AADs)).¹⁴⁰⁻¹⁴² The report of the sponsor's model was particularly useful, as they used a DES technique. The detailed critique provided by the Evidence Review Group and by the Appraisal Committee was also very helpful in identifying important factors to include in our model.

- **Ablation:** Finally, we considered an HTA review and economic evaluation that compared AAD therapy with radio frequency catheter ablation for the curative treatment of atrial fibrillation and flutter.¹⁴³

This list illustrates the wide range of models evaluating different parts of the AF service pathway, but we did not find any models that brought together all of these elements in sufficient detail to provide a platform for economic evaluation across the whole pathway.

Conceptual model development

Before constructing the computer model, a conceptual understanding and definition of the problem area was developed. This comprised two key elements: a) a model of the service pathway defined in CG36 and b) a model of the disease processes.

The design of the service pathway model began with detailed consideration of the full guideline documentation for CG36 to develop an understanding of the recommendations, the available evidence and the Guideline Development Group (GDG) rationale for decisions.¹²⁷ The NICE Quick Reference Guide (QRG) document was also useful, as this contains a set of flowcharts or other illustrations that put the recommendations together into a connected pathway.⁵⁶ For CG36, this included: an overview of the whole process from diagnosis to follow-up; strategies for cardioversion in acute and non-acute situations; a decision tree defining the criteria for selecting rate or rhythm control strategies; risk stratification and choice of drugs for prevention of stroke; and sequencing of rhythm and rate control drugs. These QRG ‘algorithms’ were developed into much more detailed and formalised flowcharts necessary to provide a foundation for the simulation model. This involved in-depth review of the full guideline and of the precise wording of the recommendations.

The conceptual service pathway model was drafted using flowcharts, which were then checked with clinical experts to identify errors or lack of clarity. Four clinicians, including a general practitioner, two cardiologists specialising in AF, and an interventional electrophysiologist provided advice. The purpose of consulting experts was to help the modellers to understand and interpret the pathway of care defined in CG36, rather than to elicit information about how services are organised in practice, or the experts’ views on how services should be organised. This process was essential to resolve some ‘gaps’ and

ambiguities in the guideline algorithms and documentation. We also sought information from the experts on sources of data to inform the model parameters.

Another essential component of the conceptual model was an understanding of the disease course; how this varies between individuals; and how it can be modified over time by interventions and events. The initial design of this disease process model for this guideline was informed by the preliminary literature review described above, and again clinical experts were invited to comment on this approach.

Boundary and scope of the model

The aim of this case study was to model the service pathway recommended in the NICE AF clinical guidelines (CG36) to estimate associated patient flows, health outcomes and costs; to assess the incremental cost-effectiveness of possible changes in the service pathway; and to estimate the value of updating selected topics within a guideline. We therefore took the scope of the NICE AF clinical guideline (CG36) as the starting point for defining the model boundaries. However, there were some differences between the scope and model boundaries which are described below.

The Clinical Guideline scope¹²⁸ included people with new onset or acute AF; chronic AF (including recurrent paroxysmal, persistent and permanent/sustained AF); co-morbidities that impact upon AF; postoperative AF; and atrial flutter that is indistinguishable from AF in terms of aim of treatment. The scope was also longitudinally broad, covering the spectrum of care for patients with all stages of the condition and associated adverse events in primary and secondary NHS healthcare settings, as well as referral to tertiary care. The guideline group considered evidence and made recommendations on:

- Identification of AF, including active case finding but not population screening;
- Investigations required to confirm diagnosis and to assess comorbidity;
- Treatment of acute-onset AF episodes with haemodynamic instability;
- Prevention and treatment of postoperative AF;
- Risk stratification and prophylactic antithrombotic treatment;
- Electrical and pharmacological interventions to promote and maintain heart rhythm;
- Pharmacological methods to control heart rate;
- Referral for specialist assessment; and

- Reviewing and monitoring of patients with AF.

The model was also broad in scope, including most of the patient groups and interventions covered in the guideline, although there were some exceptions. Due to a lack of clarity and evidence about atrial flutter this was not explicitly modelled as a separate group. We also chose not to model postoperative AF. This was a pragmatic decision, due to anticipated difficulties in reviewing a separate body of epidemiological and clinical evidence with limited time and resources. To evaluate preventive treatments for postoperative AF would also have meant introducing a very different cohort of patients into the model, including patients undergoing cardiac surgery who did not go on to develop AF.

Although we considered adopting a population approach to modelling, to reflect costs and outcomes for both prevalent and incident cases across time, this was not possible in the time available. Instead, the model took a more conventional incident cohort approach, starting with a group of individuals being tested for suspected AF, and following these individuals through until death. Information about the demographic and clinical characteristics of individuals entering the model, which governs their risks of adverse events and health outcomes, was taken from a primary care database (described below). A dataset of patients not diagnosed with AF was also obtained to allow modelling of case finding approaches and to capture consequences of false positive and false negative test results. (In the event, these data were not used as we did not identify sufficient information on the effectiveness of case finding or diagnostic test accuracy to model these questions). Extension of the model to postoperative AF would require a similar individual-patient dataset for this population, or sufficient information to generate such a dataset.

CG36 did not review evidence relating to specialist interventions to identify and correct structural heart abnormalities or electrophysiological problems, which might be the underlying cause of AF for some individuals. Evaluation of implantable devices was explicitly excluded from the scope, as was evaluation of 'novel/experimental' arrhythmia surgery. The guideline group did recommend referral to a specialist if symptoms could not be adequately controlled with conventional rate or rhythm control strategies. But they did not recommend which further treatment options should be considered for which patients. There is currently a lot of interest in various ablation techniques that are potentially curative of AF refractory to antiarrhythmic drug treatment.¹⁴³⁻¹⁴⁸ However, we did not explicitly

include this within the model as it was considered outside of the scope. The model therefore stopped, as did the guideline, at the point where patients were referred to a tertiary specialist. This issue is discussed further below.

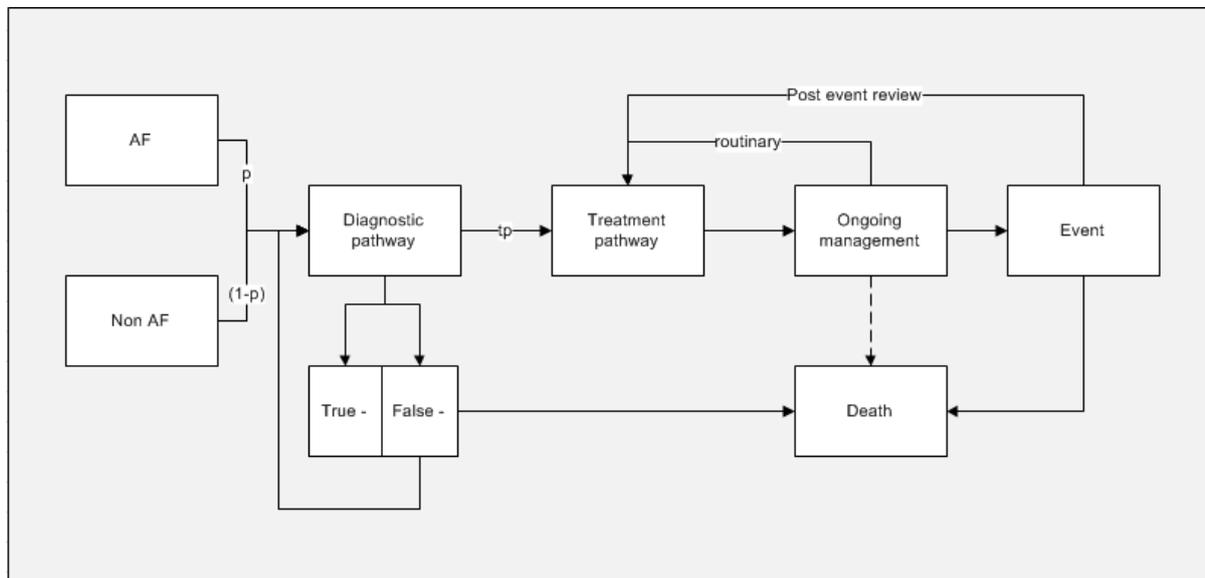
Another common boundary issue for guideline models is the evolution of adjacent and sometimes overlapping NICE guidance. During the course of this project, three NICE Technological Appraisals within the model boundary were published: two related to new oral anticoagulants, Dabigatran¹³⁷ and Rivaroxaban¹³⁸ and one to a new antiarrhythmic drug, Dronedarone¹⁴¹. As NICE Clinical Guidelines are expected to integrate current NICE Technology Appraisals unchanged, we reviewed the evidence from these published appraisals and attempted to integrate their recommendations in the CG36 pathway.

The service pathway

Outline of pathway

Figure 15 gives a broad view of the flow of patients through the service pathway.

Figure 15. Overview of AF service pathway model.



To enable evaluation of case finding and screening strategies, two cohorts of patients can be fed into the model – a cohort with AF and a cohort without AF. The proportion of patients with AF (p) can be manipulated to represent more or less targeted case finding strategies, with different rates of prevalence in the population being tested. In the analyses presented below, however, we only model results for patients presenting with ‘true’ AF.

Patients enter via the diagnostic module, where they undergo a series of tests. If the tests are negative, patients leave the diagnostic module and wait for the next event. Patients with false negative results, undiagnosed AF, are then at risk of another symptomatic AF episode or an AF-related event, such as a stroke or thromboembolism. If this event is not fatal, they will then re-present and enter the diagnostic module again. Patients with false negative results who do not experience a symptomatic episode or AF-related event wait in the model until they die from other causes.

Patients diagnosed with AF enter the treatment pathway, where they have their risk assessed and are allocated treatments based on their personal characteristics and the guideline recommendations. These treatments may include anti-thrombotic drugs, interventions to promote and maintain sinus rhythm, and/or drugs to control their heart rate. These options are discussed below.

After treatment allocation, patients enter the ongoing management module, where they wait for the next event. This can be a routine follow-up appointment, in which case they will cycle back through the treatment pathway, and possibly have their treatment reassessed. Alternatively, they may experience an event, which may include recurrence of arrhythmia, stroke or thromboembolism or an increase in another risk factor, such as onset of hypertension or diabetes. Unless the event is fatal, the patient then returns to the treatment pathway, and have their treatment reassessed.

Patients may cycle between the treatment pathway and ongoing management modules many times over their lifetime, reflecting the chronic nature of AF. The rate at which patients experience events and return for reassessment is governed by their initial characteristics on model entry, their history of events within the model, and the treatments that they receive. Patients can also leave the model at any time, due to death from non-AF related causes.

Figure 16 expands on the contents of the diagnostic and treatment pathways. This contains eight blocks related to eight main aspects of the pathway, and each linked to a chapter in CG36. Each block is also further expanded into a detailed flowchart (see Appendix 6).

Diagnosis (D)

Patients can enter the model through two routes: (a) primary care referral; and (b) emergency attendance due to an acute onset of AF. The acute onset pathway is described later. Patients presenting routinely may be symptomatic, or they may have asymptomatic AF detected incidentally – for example, by pulse palpitation in a consultation for another purpose. AF symptoms range from breathlessness or palpitations through to acute medical problems such as heart failure, stroke or thromboembolism. The precipitating trigger for an AF test is not modelled, although patients may arrive with a history of AF-related conditions and an average utility reduction is applied to reflect other symptoms.

Patients entering the model with suspected AF are referred to a specialist for an echocardiogram (ECG) (D1). AF can be missed by an ECG test, since it is often intermittent in nature (paroxysmal AF). If AF is not confirmed by the ECG and the patient is not suspected of having paroxysmal AF (D2), they will be discharged (D13). However, a negative ECG might be accompanied by suspicion of paroxysmal AF – for example, if the patient reports symptoms such as a fast heartbeat. In this case, an ambulatory ECG test might be performed, either: (a) an event recorder ECG (D8); or (b) a 24-hour ambulatory ECG monitor (D9). In general, the 24h ambulatory ECG would be used in patients with suspected asymptomatic episodes or symptomatic episodes less than 24 hours apart, whereas the event recorder ECG would be used in those with symptomatic episodes more than 24 hours apart. If an ambulatory ECG test is negative, and the doctor has a high index of suspicion, a further event recorder ECG might be requested. The model assumes that patients can receive up to three negative ambulatory ECG tests before being discharged from the system (D13).

After diagnosis of AF (by standard or ambulatory ECG), the patient might be referred for additional tests, including a Transthoracic Echocardiogram (TTE) (D4) and possibly also a Transesophageal Echocardiogram (TOE) (D6). TTE and TOE may be used to diagnose structural heart defects or to plan cardioversion. However, these treatments were not included in the model, as this was outside the scope of CG36.

The diagnostic pathway is further illustrated in the decision tree in Figure 17. This made use of data on the diagnostic accuracy of ECG from the HTA report by Hobbs and colleagues¹²⁹, although data to populate this decision tree were sparse.

Figure 16. Overview of AF service pathway model

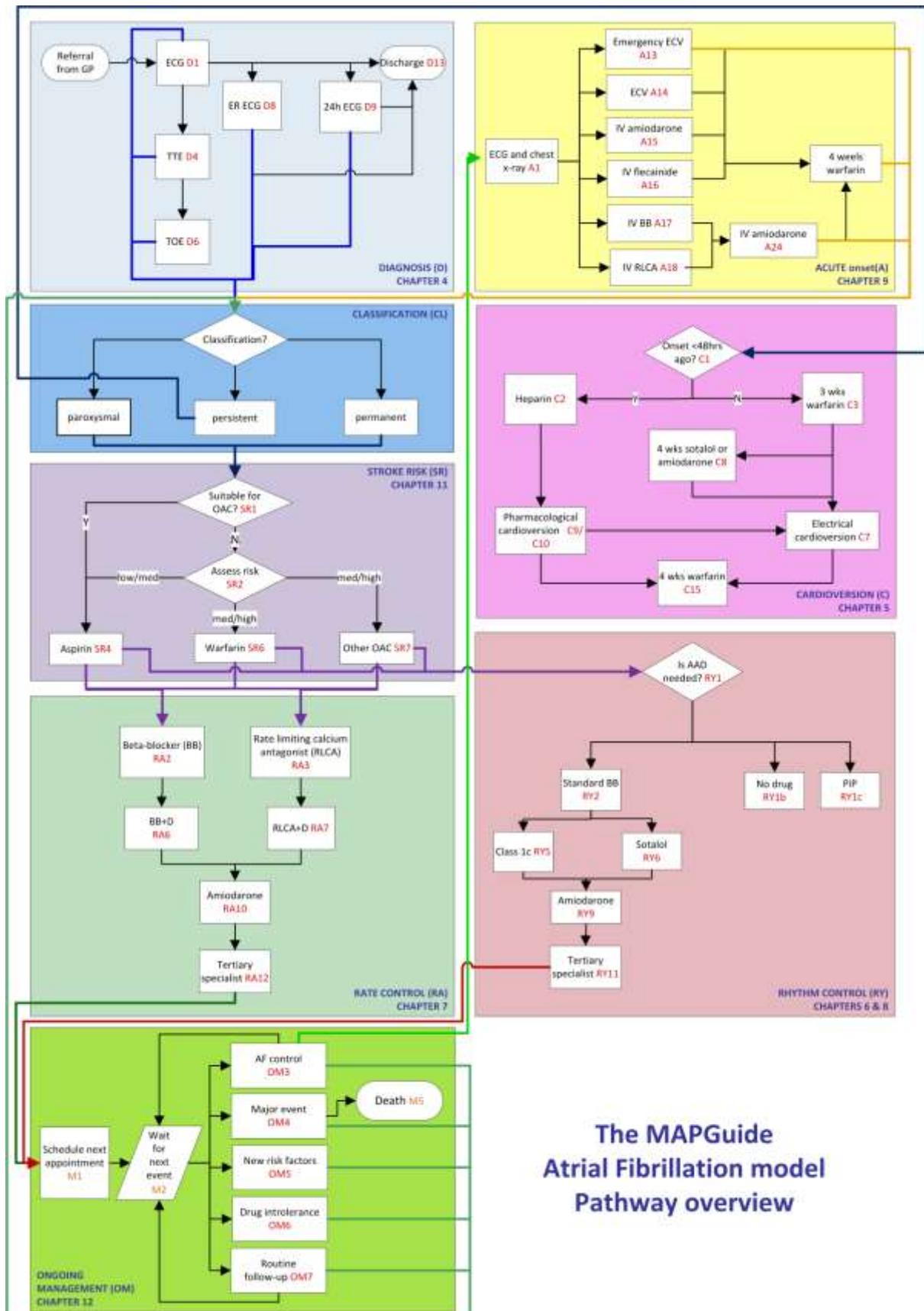
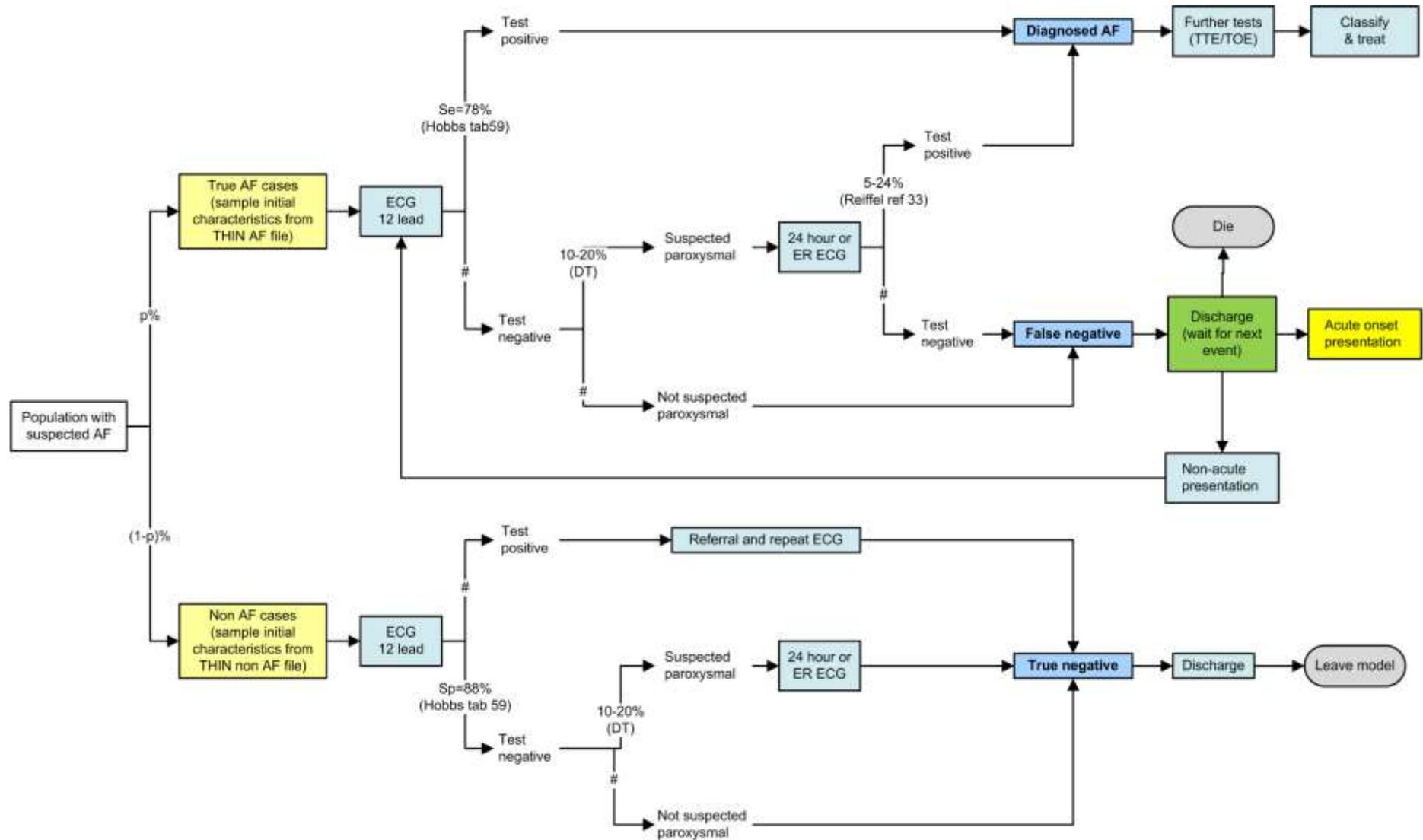


Figure 17. Decision tree showing detail of diagnostic pathway



Classification (CL)

After diagnosis, patients are classified into the three types of AF: paroxysmal (spontaneous termination < 7 days and most often < 48 hours); persistent (not self-terminating, and lasting > 7 days or prior cardioversion); or permanent (not terminated, terminated but relapsed or failed cardioversion attempt). The main significance of this classification is that it is used, along with other criteria to choose the AF treatment strategy.

- a. Patients with paroxysmal AF will usually follow a rhythm control strategy, with antiarrhythmic drugs (AADs) used to reduce the frequency of subsequent AF episodes.
- b. Patients with permanent AF will follow a rate control strategy, in which no attempt is made to regain sinus rhythm, but instead drugs are used to control heart rate and avoid symptoms and potentially dangerous tachycardia.
- c. Patients with persistent AF may follow either a rate or rhythm control strategy: CG36 defined criteria to inform this choice based on the patient's age, whether they have a history of coronary artery disease or left ventricular dysfunction, or if they are unsuitable for cardioversion or contraindicated to AADs.

Regardless of the strategy for treating AF, it is recommended that all patients should have their stroke risk assessed (SR).

Assess stroke risk (SR)

If patients are contraindicated to the oral anticoagulants (OACs) (SR1), then they are prescribed aspirin (SR4). Patients not contraindicated to an OAC (SR1) have their stroke risk assessed (SR2). There are three levels of risk defined in CG36:

- a) Low risk (age <65 with no moderate or high risk factors);
- b) Moderate risk (age \geq 65 with no high risk factors or age <75 with hypertension, diabetes or vascular disease); and
- c) High risk (previous ischaemic stroke/TIA or thromboembolic event age \geq 75 with hypertension, diabetes or vascular disease or heart failure, or impaired left ventricular function on echocardiography).

Patients at low risk are recommended for aspirin (SR4). Patients at moderate stroke risk may be treated with either aspirin (SR4) or warfarin (SR6). If patients are assessed as high risk then they will get treated either with warfarin (SR6) or dabigatran/rivaroxaban (SR7).

After stroke risk assessment, patients proceed to either rate or rhythm control treatment.

Rhythm control for paroxysmal AF (RYpx)

Patients assigned to a rhythm control strategy for paroxysmal AF (RY), might choose at first not to receive any AAD treatment (RY1b) or to use a 'pill-in-the pocket' approach (RY1c) if this is suitable. The first line of regular AAD treatment is a standard beta-blocker (RY2). After failure of treatment with a beta-blocker, the next line of treatment is either a Class 1c agent (RY5), sotalol (RY6), or amiodarone (RY9), depending on whether the patient has a history of coronary heart failure or coronary artery disease. In addition, NICE TA197¹⁴¹ recommends dronedarone as a second-line treatment for patients with additional risk factors. Once AAD treatment has failed, the guideline recommends that patients are referred to a tertiary specialist (RY11) for consideration for ablative treatment.

When patients have been allocated to an AAD treatment, they progress to ongoing management.

Cardioversion (C)

Patients with persistent AF following a rhythm control strategy will undergo a trial of cardioversion. If the onset of AF was more than 48 hours before the cardioversion (C1), electrical cardioversion (ECV) is recommended (C7), preceded by three weeks of warfarin and/or TOE guided ECV to reduce the risk of thromboembolism (C3). Patients with a high risk of cardioversion failure will also receive 4 weeks of sotalol or amiodarone (C8) before ECV. Patients with AF onset less than 48 hours ago would benefit from speedier treatment, so prophylaxis with heparin administered by injection (C2) is used prior to cardioversion. In these patients, the guideline recommends use of either electrical or pharmacological cardioversion (PCV). Patients with structural heart disease (coronary artery disease or left ventricular dysfunction) undergoing PCV will be treated with intravenous amiodarone, otherwise a Class 1c agent is recommended. If cardioversion (PCV or ECV) is not successful, the procedure can be repeated. The model assumes a maximum of two cardioversion attempts.

After an attempt at cardioversion, patients will have their stroke risk assessed (SR), before proceeding to the rate or rhythm control modules.

Rhythm control for persistent AF (RYps)

The sequencing of antiarrhythmic drugs for patients with persistent AF following a rhythm control strategy is similar to that for patients with paroxysmal AF, except that a pill-in-the-pocket approach or no treatment is not usually considered appropriate. After initiation of an AAD, patients enter the ongoing management (OM) strategy.

Rate control for persistent and permanent AF (RA)

The rate control strategy contains three lines of drug treatment, followed by referral to a tertiary specialist if the heart rate remains uncontrolled (>80bpm). The first line is a rate-limiting calcium antagonist (RA3) if heart rate control during exercise is required, or otherwise beta-blockers (RA2). If these treatments are unsuccessful at controlling the heart rate, digoxin is added (RA6 & RA7), followed by amiodarone (RA10).

Acute onset AF (A)

Patients presenting with an acute arrhythmia associated with haemodynamic instability will first receive an ECG, chest x-ray and check of electrolytes (A1) to establish the cause of the hemodynamic instability if possible. If the situation is life threatening, an emergency ECV will be performed (A13). If the haemodynamic instability is not life-threatening, patients not already taking anticoagulants will be given heparin (if not contraindicated) before proceeding to treatment.

For patients known to have permanent AF, then urgent treatment with an intravenous rate control drug will be used to reduce the heart rate. This will usually be either a beta-blocker (A17) or rate-limiting calcium antagonist (A18), although amiodarone (A24) may also be tried. If the AF is not known to be permanent, then urgent rhythm control with cardioversion will be tried (A9). ECV (A14) is recommended in this context, although PCV (A15 & A16) may be used if there is a delay in organising ECV. For PCV, the guideline recommends intravenous flecainide if the patient is known to have Wolff Parkinson-White syndrome, or intravenous amiodarone otherwise.

After treatment, patients with onset of AF less than 48 hours previously or at high risk of recurrence will be offered 4 weeks of warfarin, before being routed to further treatment.

Ongoing management (OM)

On entering the ongoing management module, an appointment with the GP or specialist (OM1) is scheduled, according to the recommended follow-up frequencies from the guideline (CG36 Chapter 12). Patients then wait (OM2) until the next one of five types of event occurs:

(a) **Loss of AF control** (OM3). For patients on a rhythm control strategy, this will be an AF recurrence (loss of sinus rhythm). A recurrence may be 'undocumented', which is not sufficiently serious to trigger a consultation and the patient continues to wait until the next event, or 'documented' which causes the patient to seek medical attention. Documented recurrences can be acute, in which case the patient is routed to the acute onset module (A). Otherwise, patients are routed to classification (CL) to be allocated to the appropriate treatment strategy. In patients on a rate control strategy, loss of control is defined as having a resting heart rate above 80 beats per minute, which may be of acute onset (route to module A) or non-acute (route to CL).

(b) **Major adverse event** (OM4). Events included in the model are thromboembolic events (ischemic stroke, TIA or other) and bleeds (haemorrhagic stroke or major bleed). These events may be fatal (M5). If the patient survives, they will be routed to the classification module (CL), where their treatment will be reassessed.

(c) **New risk factor** (OM5). The onset of new risk factors, such as hypertension, diabetes or passing an age threshold, can have two effects. Firstly, it can increase individuals' risk of major events, reducing the time to their next major event within the ongoing management module. Secondly, additional risk factors might trigger a change in treatment, as patients meet criteria that they would have previously failed. In this case patients are routed back to classification (CL) to have their treatment adjusted.

(d) **Drug withdrawal** (OM6). Patients might stop taking a drug, either due to an adverse effect or for some other reason. After a drug withdrawal, patients are sent to classification (CL), and will pass again through the pathway to have alternative treatment considered.

(e) **Routine follow-up** (OM7). It is assumed that previously undocumented AF recurrence will be detected at this time, when patients are asked about symptoms and have further tests. In such cases, patients are returned to classification (CL) and have their treatment

reconsidered. Otherwise, patients have their next routine visit scheduled, and then wait for the next event (M2).

The disease process model

The above rules were based a model of the risks associated with AF, as illustrated in Figure 18. This is built around the five types of outcome shown in the column on the far right:

- **Loss of AF control.** This was defined as loss of sinus rhythm for patients being treated with a rhythm control strategy (paroxysmal and some persistent). In patients with paroxysmal AF, this loss of sinus rhythm could be documented or undocumented, depending on whether the symptoms were sufficient for the patient to seek medical attention. For patients being treated under a rate control strategy, AF control was defined as maintenance of a resting heart rate below 80 beats per minute.
- **Thromboembolisms.** These were defined to include ischaemic strokes, Transient Ischaemic Attacks (TIA) and other major thromboembolic events. The risk of thromboembolism is greatly increased with AF, and so it is an important outcome to include within the model.
- **Bleeds.** Including haemorrhagic strokes and other major bleeds. These events are included as an outcome, because drug treatment to prevent thromboembolism increases the risk of bleeding.
- **Other related risks.** The incidences of several other conditions (hypertension, diabetes, coronary heart disease and heart failure) were modelled as risk factors for the above directly AF-related outcomes.
- **Death.** Mortality unrelated to AF was modelled independently of the other risk factors (other than age and sex). Mortality related to AF was modelled by applying case-fatality rates to acute-onset arrhythmias, thromboembolic events and bleeds.

Loss of AF control, thromboembolisms and bleeds impact directly on health status (and hence on QALYs) in two ways: they can be fatal; and in patients who survive the event, they can reduce quality of life (utility). Patients who survive also incur additional treatment costs. We did not include costs or QALY losses for other conditions modelled as risk factors (hypertension, diabetes, CHD and heart failure). We assumed that ischaemic and haemorrhagic strokes would have a lasting impact on utility and healthcare costs, due to the

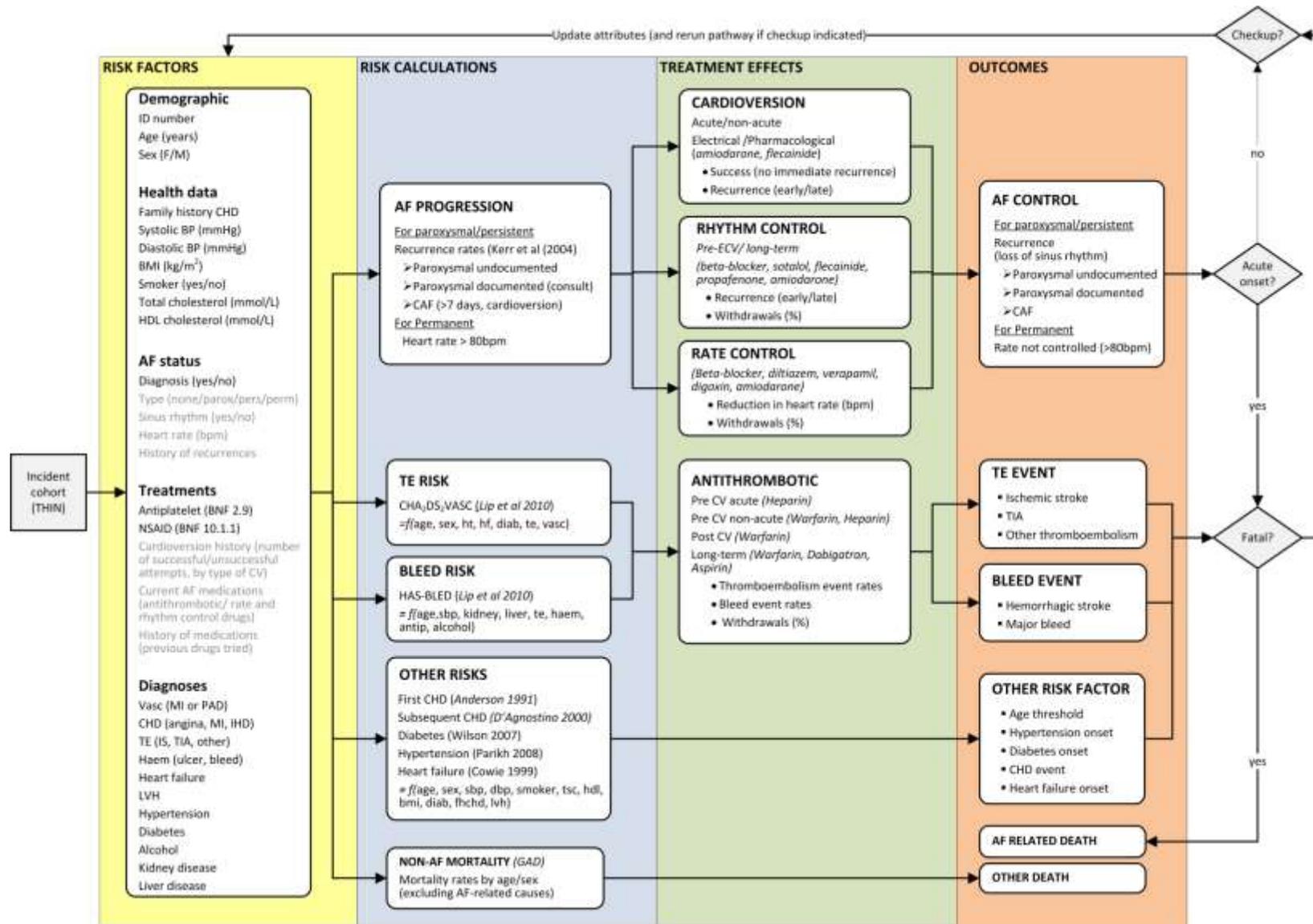
high potential for disability. Other events were assumed to have more transient consequences, incurring costs and utility decrements over a short initial period.

The risk equations or algorithms listed in the second column of Figure 18 were used to calculate individuals' risks of the included outcomes in the absence of treatment. There are five main classes of risk calculation used in the model, based on the five types of outcome. The risks of loss of AF control and progression between the types of AF (paroxysmal, persistent and permanent) were defined according a model that we developed. The risks of thromboembolisms and bleeds were defined by published risk algorithms for patients with AF: CHA₂DS₂-VASc and HAS-BLED respectively.¹⁴⁹⁻¹⁵¹ Rates of incidence for the other risk factors were also based on published sources: Framingham equations for CHD, type 2 diabetes and hypertension¹⁵²⁻¹⁵⁴, and simple age and sex based incidence from a cohort study for heart failure¹⁵⁵. Mortality rates from non-AF related were based on national life table data.⁸⁶

The inputs required for these risk calculations define the set of individual risk factor information that is required for the model – listed in the first column of Figure 18. The risk factors in bold were defined in advance of model entry, as variables from our individual-patient dataset from The Health Improvement Network (THIN). The factors in grey are assigned as patients move through the simulation model.

Finally, the third column in Figure 18 lists the treatment effects that are used to modify individuals' baseline risks in accordance with any treatments that they receive. Treatments are grouped into four classes, defined by their major outcome targets; cardioversion (aim to regain sinus rhythm), rhythm control drugs (aim to prevent AF recurrence); rate control (aim to achieve control of heart rate); and antithrombotic (aim to reduce the risk of thromboembolism, while minimising impacts on bleeding). In addition, a withdrawal rate was defined for each drug.

Figure 18. Atrial fibrillation disease process model



AF progression and control

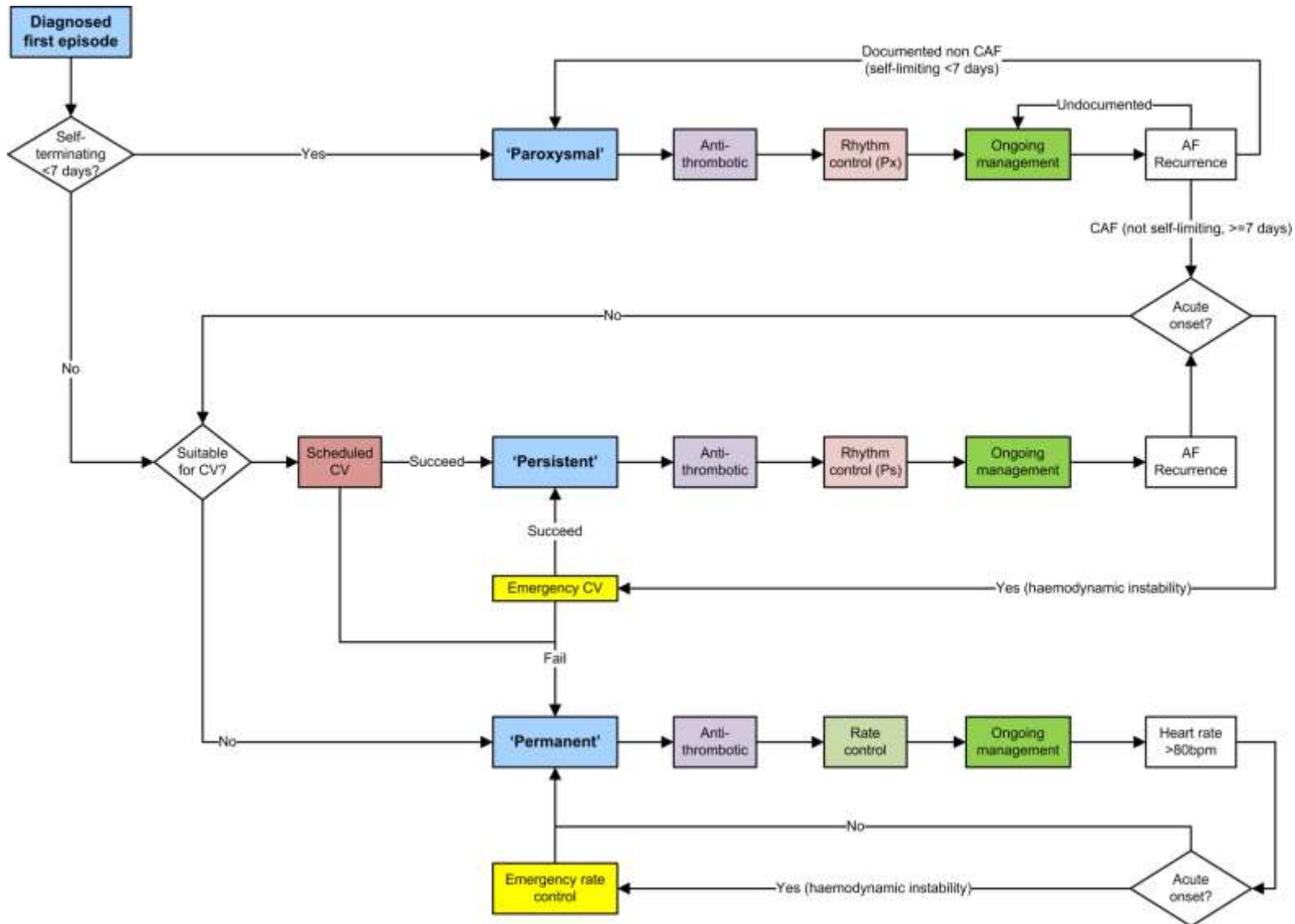
The process by which patients pass through the different types of AF is illustrated in Figure 19. If the first diagnosed episode terminates without intervention within seven days, patients are classified as having 'paroxysmal' AF. They pass through the service pathway, as described above, are prescribed appropriate antithrombotic and antiarrhythmic therapy, and move into the ongoing management module. If they have an AF recurrence, this will be one of three types: an 'undocumented' recurrence for which they do not seek medical attention and remain in ongoing management; a documented recurrence that is self-terminating within seven days but leads to a reassessment of their antithrombotic and antiarrhythmic medication; or onset of 'chronic AF' (CAF) that does not self-terminate within seven days. The latter defines a transition from paroxysmal to persistent AF.

Patients with an episode of AF that is not self-terminating in 7 days (either a first episode or a CAF recurrence of paroxysmal AF) are considered for cardioversion. If they are suitable for this procedure, it is scheduled, and if successful the patient is classified as having 'persistent' AF. Patients with persistent AF are prescribed appropriate antithrombotic and antiarrhythmic therapy, before going to ongoing management. If they have a recurrence, it is assumed that they would require cardioversion to move back into sinus rhythm. This is a simplification, as in reality patients with persistent AF may also experience paroxysmal episodes.

A patient with AF that has not terminated within seven days is not suitable for cardioversion, for whom cardioversion has failed, is classified as having 'permanent' AF. Patients with permanent AF are given appropriate antithrombotic and rate control drugs, before going to ongoing management. If the rate control drugs are insufficient to bring their resting heart rate below the threshold of 80 beats per minute, their treatment will be reassessed at their next scheduled appointment.

When patients experience an AF recurrence (paroxysmal or persistent) or uncontrolled heart rate (permanent), this may be of acute onset necessitating an emergency cardioversion or rate control intervention.

Figure 19. Model of AF progression and control



Risk of thromboembolism (CHA₂DS₂VASc)

Atrial fibrillation is associated with a substantial risk of stroke and other thromboembolism. This risk is not homogenous and various risk factors have been identified that are predictive of stroke in AF.¹⁵⁶ These risk factors have been formulated into various stroke risk stratification schemes.^{150;156-161}

A well-known and simple risk assessment scheme is the CHADS₂ score.¹⁶² This evolved from the AF investigators and Stroke Prevention in Atrial Fibrillation (SPAF) Investigators criteria¹⁶³, and is based on a point system in which two points are assigned for a history of stroke or TIA and one point each is assigned for age >75 years, a history of hypertension, diabetes, or recent cardiac failure. The CHA₂DS₂VASc extends the CHADS₂ scheme by adding vascular disease (myocardial infarction, peripheral arterial disease or aortic plaque) as a risk factor.¹⁵⁰ The score is calculated by adding one for each risk factor, and an additional point each for age >75 years and prior stroke.

Risk of bleeding (HAS-BLED)

Various risk stratification systems have been proposed for the assessment of the risk of bleeding.^{149;164;165} Individuals' risk of bleeding in the model is assigned on the basis of the HAS-BLED risk algorithm.¹⁵¹ This was developed from a cohort study of 3,978 European subjects with AF from the EuroHeart Survey. The HAS-BLED score is calculated by adding one for each of the following risk factors: hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile INR, elderly (>65), drugs that increase the risk of bleeding (e.g. NSAIDs or aspirin), alcohol (8 units per week or more). Coefficient calculations exclude data on labile INR as this was not available for the cohort. We also excluded antiplatelet therapy when calculating individual HAS-BLED scores in the model, to avoid double-counting the effect of aspirin on bleeding rates (as we applied a relative risks of bleeding with aspirin and oral anticoagulants).

Other related risks

The risks of new-onset Coronary Heart Disease (CHD), hypertension and diabetes were calculated using multivariate risk equations estimated from the Framingham cohort study.¹⁵²⁻¹⁵⁴ Finally, the risk of onset of heart failure was estimated based on age/sex

specific rates from a general population cohort.¹⁵⁵ These sources are not ideal for our purpose, since they are estimated from a general population cohort, rather than from people with AF.

Model design

DES model

The DES model combines the conceptual service pathway (Figure 16) and the disease process model (Figure 18) into a single dynamic incident cohort model that incorporates time. The patients are modelled as individual entities. Each has labels attached, which record their personal characteristics, including the risk factors listed in the first column of Figure 18, as well as a record of their treatment and event history that accumulates within the model. The values assigned to each label may change as the model runs.

The patients travel through the model accruing costs and QALYS as they receive treatment and experience events. The patient's route may be determined by the values stored in their labels (criteria-based routing), or by sampling from a defined distribution (probability-based routing). For instance, in the classification module the decision to adopt a rate or rhythm control strategy for a particular patient is informed by the contents of the label that records whether they have paroxysmal, persistent or permanent AF, whereas in the Diagnosis module, the choice between a 24 hour or event recorder ECG is randomly decided according to defined probabilities. In this way, each patient's route through the model is tailored to their individual characteristics, but also depends to some extent on chance.

Selection of the patient cohort

The model contains information on 12,776 patients with newly diagnosed AF, drawn from The Health Improvement Network (THIN) primary care database (see data sources section below). The database contains the characteristics listed in Table 22 for each patient.

When the model is run, it is possible to randomly select patients from a sub-set based on their initial characteristics. For example, it would be possible to select a group of patients within a specific age range who have a history of hypertension. Sampling is from the list of eligible patients is random 'with replacement'. The cohort of patients arrives in the system at time zero, and the model runs until all patients from the cohort have died.

Table 22. Risk factors drawn from THIN data

	Risk factors	Values	Simul8 label
<i>id</i>	Unique identifier for patients	Integer	
<i>sex</i>	Sex	0=male, 1=female	i_Sex
<i>age</i>	Age	years	i_CurrentAge
<i>diagaf</i>	Incident AF cases	0=no, 1=yes	i_DrawnFromKnownAFpatientsDatabase
<i>fhchd</i>	Family history of diabetes/hypertension	0=no, 1=yes	h_ParentalDiabetes, h_ParentalHypertension
<i>sbp</i>	Systolic blood pressure	mmHg	i_BloodPressureDiastolic
<i>dbp</i>	Diastolic blood pressure	mmHg	i_BloodPressureSystolic
<i>bmi</i>	Body mass index (height/weight squared)	kg/m2	i_BMI
<i>smoker</i>	Current smoker	0=no, 1=yes	h_smoker
<i>tsc</i>	Total serum cholesterol	mmol/L	i_Cholesterol_Total
<i>hdl</i>	HDL cholesterol	mmol/L	i_Cholesterol_HDL
<i>antipl</i>	Antiplatelet drugs (BNF 2.9)	0=no, 1=yes	m_AntiplateletDrugs
<i>nsaid</i>	Non-steroidal anti-inflammatory drugs (BNF 10.1.1)	0=no, 1=yes	m_NSAID
<i>vasc</i>	Vascular disease: MI (I21, I252) or PAD (I70-73)	0=no, 1=yes	h_VascularDisease_MI_PAD
<i>chd</i>	CHD: Angina, MI, coronary insufficiency	0=no, 1=yes	h_CHD_Angina,_MI_coronary insufficiency
<i>te</i>	Thromboembolism: IS (I63-4), TIA (G45) or Other TE (I74,I26)	0=no, 1=yes	h_ThromboembolicEvent
<i>haem</i>	Bleed: Intracranial (I160-2) or other major bleed (I850, I983, K25-28 (0-2,4-6), K625, K922, D629)	0=no, 1=yes	h_BleedingHistory
<i>hf</i>	Heart Failure: CHF/LVD (I50)	0=no, 1=yes	h_HeartFailure_LVD_CHF
<i>lvh</i>	Left Ventricular Hypertrophy	0=no, 1=yes	h_LVHypertrophy
<i>ht</i>	Hypertension (I10-15) - CHA ₂ DS ₂ -VASc definition	0=no, 1=yes	h_Hypertension
<i>diab</i>	Diabetes (E10-14)	0=no, 1=yes	h_Diabetes
<i>alcohol</i>	Alcoholic disease	0=no, 1=yes	h_AlcoholConsumption
<i>kidney</i>	Renal disease (N17-19, transplant or dialysis)	0=no, 1=yes	h_RenalFunction
<i>liver</i>	Liver disease (K70-77, transplant or resection)	0=no, 1=yes	h_LiverFunction

Set up attributes

When the sampled patients arrive in the system, information about their starting state is read from the database and copied to the appropriate labels. Patients are assigned a label to specify that they have AF that is currently undiagnosed (CurrentAFstate=5). Each patient is also assigned a label (Non_AFdeathAge) which specifies their age of death, unless they die from an AF-related cause prior to this. National life table data were used to generate probability distributions for life expectancy, based on the patient's initial age and sex. Patients' starting utility is assigned (CurrentUtility) based on their age and sex. This is obtained from a look-up table, so all patients of the same age and sex will have the same starting utility. Labels are also set to indicate that patients are not initially taking any medication. Each patient has a random number (U) assigned for each of the events to be modelled (e.g. U_bleed, U_diabetes). These are used when calculating, and updating, the time to each event. Finally all patients arriving have a label i_NextEvent initialised to "First Event". This next event label is used to control the routing of the patient through the on-going management section of the model.

Diagnosis

The diagnostic section of the model contains more probability-based decisions than the rest of the model. This reflects the difficulty that we experienced in obtaining data for this module. The intermittent nature of AF makes it impossible to establish false negative rates for the diagnostic tests, and there is no 'gold standard' for assessment of diagnostic accuracy. The model does include the facility to add patients without AF, and to include false positive test results for these patients, incurring unnecessary expenditure. However, this was not applied in the analyses presented below.

Diagnosis is one of the two sections of the pathway in which time elapses as the patient progresses. Patients arrive and have to wait a number of days for their ECG based on a defined distribution (mean 14 days, standard deviation 3 days). Costs associated with ECG are added to the patient's tally of costs. For patients with AF, there are four possible routes after ECG: a) the patient has confirmed AF and there is no need for more tests; b) the patient has confirmed AF but needs a TTE/ TOE

examinations to assess underlying physical problems; c) the patient's ECG was negative but there is still a suspicion of AF so they will receive either 24 hour ambulatory ECG or an event recorder ECG; or d) the patient has a negative ECG and there is no suspicion of AF (false negatives). Patients who required an ambulatory ECG are randomly allocated to either a 24 hour or event recorder test. Those patients with negative results may then receive an event recorder ECG or be incorrectly discharged (false negatives), again this is randomly allocated.

After a positive diagnosis, patients are randomly allocated to have paroxysmal or persistent AF, in accordance with a defined probability, and their CurrentAFstatus label is updated to 1 (paroxysmal), or 2 (persistent). It is assumed that no patients present for the first time with permanent AF.

False negatives

The patients who receive a false negative diagnosis remain in the system to ensure a penalty for missing cases is incorporated. These patients loop through a similar (but reduced version) of the ongoing management module (described below). These untreated AF patients have their risk factors and characteristics updated in the same way as treated patients, and incur the costs and consequences of any major events that occur. If these patients experience an arrhythmic event then depending on the severity they will either re-present for an ECG – effectively re-starting this process after a delay, or they will present at the Accident and Emergency department for cardioversion. As these patients will not be taking medication for their AF, they do not receive any protective benefit from antithrombotic, rhythm or rate control drugs.

Classification

All patients arriving in the classification section of the model will have the CurrentAFstate label set to 1 (paroxysmal), 2 (persistent), or 3 (permanent). Only those who have cycled through the system at least once can be classified as having permanent AF. Patients with paroxysmal AF are assigned to a rhythm control treatment strategy; patients with permanent AF are assigned to rate control treatment strategy; and patients with persistent AF are assigned to either a rate

control or rhythm control strategy depending on their attributes in accordance with the guidelines. Patients have their TreatmentOption label updated to reflect the strategy that is adopted, and are directed to cardioversion or stroke risk classification as appropriate.

Cardioversion

Patients with persistent AF who are assigned to a rhythm control strategy, and who are not in sinus rhythm when they pass through the classification process receive non-emergency cardioversion. The number of cardioversion attempts per episode is limited, currently set to a maximum of two attempts, though this parameter may be changed. If cardioversion is unsuccessful, then the patient is considered to be in permanent AF, their i_CurrentAFstatus label is updated, and they are assigned to a rate control strategy. If cardioversion is successful then they remain on a rhythm control strategy and remain labelled as having persistent AF. Following cardioversion, patients are routed to have their stroke risk assessed.

Stroke risk assessment and antithrombotic therapy

When patients enter this module, their stroke risk scores used for treatment allocation are recalculated (CHADS₂, CHA₂DS₂VASc, or the NICE clinical guideline stroke risk algorithm). If this has increased since their previous stroke risk assessment then the anti-coagulation therapy is also re-assessed, otherwise they remain on their current therapy.

The NICE criteria for assignment of anticoagulation depend on age and various other risk factors (CG36, TA249 and TA256). In many cases patients are eligible for more than one drug, so some assumptions are needed to determine the split of patients between these options. An example of the percentage split between assumptions used for the 65-74 year old age group is shown in Table 23. This shows the assumed percentage split between the eligible drugs for each risk group. If a patient is contraindicated to a drug or has previously withdrawn from taking it, the percentages are revised to ensure that the total equals 100% across all eligible options. The patient is then randomly allocated a suitable anticoagulant based on these revised percentages.

Table 23 Choice of antithrombotic drugs by risk factors: age 65-74

	Aspirin	Warfarin	Dabigatran	Rivaroxaban
0	50%	50%	-	-
1	33%	33%	17%	17%
2	-	50%	25%	25%
>2	-	50%	25%	25%
Previous stroke	-	50%	25%	25%
LVD or CHF	-	50%	25%	25%

Rate and rhythm control

Each patient has two labels in which their current medication is recorded, one for antithrombotic drugs and one for rate or rhythm control drugs. These record the row number associated with the particular drug, as shown in Table 24. The order of the rows for all the tables associated with the drugs (such as utilities, costs etc.) is the same throughout the model. This allows the addition of other drugs at a later stage.

The selection of the next line of rate or rhythm control treatment is dictated partially by the guideline and, where there is a choice of medication, by a similar random process as described above for anticoagulants. There are three tables – one each for paroxysmal, persistent (rhythm control), and rate control - that record the recommended sequencing of medications. These tables contain 21 medications on both the horizontal and vertical axes. The cells within the tables detail the percentage chance of changing from any given drug (vertical axis) to another (horizontal axis). Similarly to the anticoagulants, the percentages are read from the table and adjusted to take account of any existing contraindications for the patient, before patients are randomly assigned to one of the remaining drugs for which they are eligible. On a change of medication, the previous drug is marked as contraindicated to prevent selection again in the future

Table 24. Drugs currently included in the model

2	First visit (medication not yet considered)
3	No drug treatment prescribed
4	Aspirin
5	Warfarin
6	Heparin
7	Dabigatran
8	Rivaroxaban
9	Pill-in-pocket
10	Standard beta-blocker (rhythm)
11	Beta-blockers (rate)
12	Beta-blockers (rate)+ digoxin
13	Calcium-channel antagonist
14	Calcium-channel antagonist + digoxin
15	Digoxin
16	Class 1c drug
17	Sotalol
18	Dronedarone
19	Amiodarone
20	Intravenous beta-blocker
21	Intravenous calcium-channel antagonist
22	Intravenous amiodarone

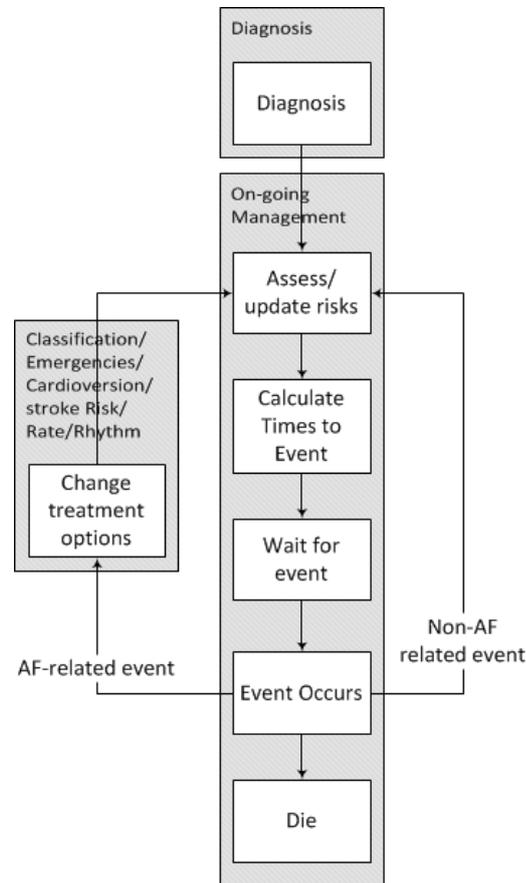
Once a patient has been prescribed a new drug they are routed to the ongoing management section of the model. Patients may pass through the rate and rhythm control sections of the model many times, as their AF progresses, or if they withdraw from a drug. A similar process of drug selection is followed for each line of treatment. Once patients have exhausted all lines of treatment they are referred to the tertiary specialist for consideration for an interventional procedure. We assume that they will continue to take the final line of treatment following referral.

Ongoing management

Patients travel through the model and after being diagnosed and receiving their first line treatment options, arrive in the ongoing management section of the model (see Figure 20). This section is the main driver behind the model. When patients enter, all of their risk scores and risk values are updated based on their current characteristics and any treatment effects. From these updated risks, a 'time-to-event' is calculated for each event of interest. These times, and the time to non-AF death, are compared to find the minimum, and this event is designated as the next event. The patient then waits until the time designated for this event occurs. This means that each

patient moves through the model in time increments defined by the events that they experience, as opposed to all moving at pre-defined time intervals.

Figure 20. Illustration of ongoing management process



If the event is not related to AF, and would not cause a change in rate or rhythm control treatment, the patient’s anti-thrombotic risk is re-assessed and changed if appropriate. The patient’s risk factors are updated, the effects of the new anticoagulant applied, and the ‘time-to-event’ recalculated. If, however, the patient’s AF has progressed, they are routed back through the classification module, from where they may change from rhythm control to rate control, be referred for cardioversion, and/or move to the next line of treatment. Following this, they return to the ongoing management, where the process repeats.

Modelling risk

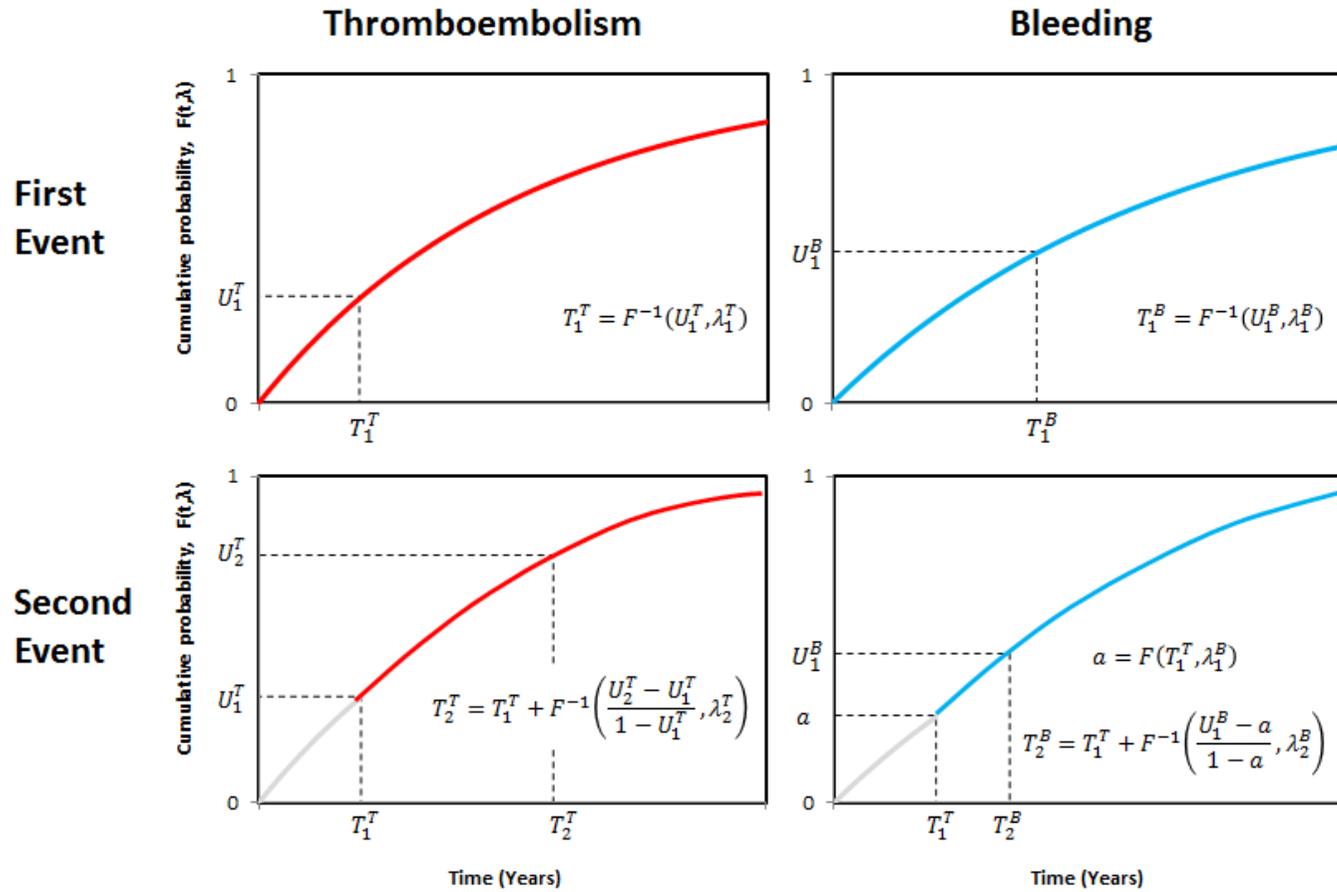
The chance of not having a particular event until a time t , given an adjusted hazard rate of λ is modelled by an exponential survival function ($F(t; \lambda) = 1 - e^{-\lambda t}$). This assumes that the hazard remains constant over the period of time modelled.

However, as patients move through the model, their risk of a particular type of event may change in response to other associated risks, or as they age. We modelled this using a piecemeal exponential distribution, in which the hazard changes at defined points in time (when an event has occurred), but is constant in between events.

The times at which events occur are determined by random numbers. At model entry, each patient is assigned a random number for each of the main types of event in the model. This could be considered as a proxy for unknown factors that influence a patient's propensity for that particular type of event. Thromboembolic events (TE events), bleeds and AF recurrences are composite events, each consisting of a number of sub-types of event. TE events comprise Ischaemic Strokes, Transient Ischaemic Attacks (TIA) and other thromboembolic events, bleeds comprise haemorrhagic strokes and other major bleeds, and recurrence events comprise undocumented events, self-terminating events, non-terminating events requiring cardioversion, and acute arrhythmic events requiring emergency cardioversion. For each composite event, patients are assigned another random number, which determines which sub-type of event will occur next. This approach ensures that related groups of events are not treated as independent.⁶¹

Once an event has occurred, care is needed in adjusting the times for competing events to avoid counterintuitive effects. The process for sampling time to event is illustrated in Figure 21. For simplicity, suppose initially that there are only two types of event, thromboembolisms and bleeds. On entry to the model, an individual is assigned two random numbers between zero and one: one for thromboembolisms U_1^T and one for bleeding U_1^B . The person's starting attributes are used to calculate CHA₂DS₂VASC and HAS-BLED and initial hazard rates for thromboembolism λ_1^T and bleeding λ_1^B are estimated.

Figure 21. Illustration of the sampling process for the AF model



The times to events are calculated using the inverse of the exponential survival function: $T_1^k = F^{-1}(U_1^k, \lambda_1^k) = -\ln(1 - U_1^k)/\lambda_1^k$, where $k = T, B$. Suppose that $T_1^T = 10$ and $T_1^B = 20$, so that a thromboembolism is the first event to occur. At year 10, having had one thromboembolism, the person is now at higher risk of a second thromboembolism and also at higher risk of a bleed. The CHA₂DS₂VASC and HAS-BLED scores are updated, the patient is assessed for any changes in treatment and revised estimates of the hazard rates: λ_2^T and λ_2^B are obtained. The time of the next thromboembolism T_2^T is calculated as before, using a new random number U_2^T . However, if we were to do the same for bleeding, there is a chance that we could draw a random number such that the first bleed would occur later than had been originally expected: $T_2^B > T_1^B$. This would be counterintuitive, as the time to a bleed would appear to have been increased by the occurrence of a thromboembolism. To avoid this, instead of drawing a new random number for bleeding, the original number U_1^B is adjusted to reflect the remaining probability of a bleed, conditional on no bleed having occurred up to time T_1^T : $U_2^B = (U_1^B - a)/(1 - a)$, where $a = F(T_1^T, \lambda_1^B)$. The time of the next bleed is then calculated as $T_2^B = T_1^T + F^{-1}(U_2^B, \lambda_2^B)$. Thus the random number used to calculate the time to an event is only re-sampled when that particular type of event has occurred. This procedure is easily extended to include other types and subtypes of events, and also age thresholds that increase estimated risks or trigger new treatments.

Calculating costs and QALYs

For each outcome event there may be short and long-term consequences, affecting costs and QALYs. For instance, an ischaemic stroke incurs a mean cost of £11,646 associated with the initial hospitalisation and treatment during the first 90 days, followed by an ongoing cost of £22.61 per day for continuing health and social care. The model assumes that a patient experiencing the event incurs the initial cost at the time of the event (regardless of whether they survive for the whole 90 days), and schedules the daily cost to start after the initial period, if they survive that long. The ongoing cost continues for the remaining lifetime of the patient.

A similar method is used to apply short and long-term modifications to patient's utility values. A short-term utility multiplier is applied during a defined initial period

after the event, followed by a long-term multiplier. The adjusted utilities are used to estimate patients' QALY for as long as they survive. The duration of the initial period can differ between events, and between the costs and utilities. Some types of event are assumed not to incur any lasting cost or utility effect after the initial period.

Costs and QALYS are updated every time the patient passes through a section of the pathway where a cost is applied, as well as whenever a change in treatment occurs. In addition, if the time to the next event is more than the 'Frequency of Update' variable (currently 90 days) then an update of costs and QALYS is scheduled. Costs and QALYS are discounted using a continuous time approach to the time of model entry for each patient.

Data sources

A list of all parameters needed to populate the model was compiled. Potentially relevant sources of information to define these parameters were first identified from the searches conducted for CG36. This provided a base of evidence of clinical effectiveness, which we knew had been identified through systematic searches and quality appraised. In extracting this evidence, we paid particular attention to the commentary on the quality and interpretation of the evidence base in the 'from evidence to recommendations' sections of the full guideline. Clinical papers that had informed GDG decisions were obtained and relevant data was extracted.

However, a recurring problem with CG36 as a source of data for modelling was the lack of meta-analysis in this document. When multiple studies related to a question were identified, the results were presented in a narrative fashion, with no attempt to statistically pool the available data. We were also aware that for some interventions in particular, the evidence base reported in CG36 was significantly out of date. It was not possible to conduct our own systematic reviews within the resources available, so where necessary we relied on other published systematic reviews and meta-analyses. Reviews conducted for NICE Technology Appraisals (TA) were prioritised as sources of evidence for the model, as we knew that they will have undergone rigorous review and public consultation. Where NICE TA reviews were not available, we sought to identify information from other HTA reports, Cochrane reviews or other high-quality sources. However, it should be emphasised that

effectiveness estimates were not all based on up-to-date systematic reviews, and that the results should therefore not be used to inform clinical decisions.

In addition to evidence about clinical effectiveness, we needed data to inform other model parameters: including background rates of the adverse events defined in the disease process model, utilities and costs. The other major data requirement for the model was information to define a representative cohort of patients to feed into the model. Individual patient data was obtained for a cohort of patients newly diagnosed with AF from a primary care database. These data are described below.

Individual-patient dataset

The Health Improvement Network (THIN) is a research database comprising anonymised patient records uploaded from primary care information systems. THIN data collection began in 2003, and currently contains data from 479 practices with a total of 9.1 million patients. We obtained an extract of data for patients registered on an index date (1 May 2008), aged 30 or older, without any record of an AF diagnosis during the two year period prior to the index date. Individuals within this group with a record of an AF diagnosis during the two year period after the index date were then identified; defining an incident cohort of 12,776 patients that was used to populate the model. Demographic and medical information was collated for these patients for the two years before and two years after the index date (from 1 May 2006 to 30 April 2010), although some patients left the system during the two years after the index date. The list of patient-level variables used in the model is shown in Table 22 above.

Some variables were recorded more than once over the four year follow-up period: for example, blood pressure and lipid levels had often been measured several times during this time. In such cases, we took the average of the three readings closest to the date of diagnosis. Some individuals did not have a record of blood pressure, BMI, lipids (total or HDL cholesterol) or smoking status. These missing data were imputed using a multivariate regression approach, to provide a full dataset for use in the model. Before imputation, the incident cases of AF were on average 73.6 years old, and 47% female. Around 5% had a family history of CHD. Their average blood pressure was 78/137 mmHg, their average BMI was 28.5kg/m² and 12% were

current smokers. In terms of medication, around 40% were on antiplatelet or lipid lowering medication, and 65% were on antihypertensive medication. 21% had a history of haemorrhage, such as an ulcer or bleed. Imputed values for missing data items had very similar means and standard deviations to those in the non-imputed data.

Risks of bleeding and thromboembolism

The model uses data from the Swedish AF cohort study to estimate incidence rates for thromboembolic and haemorrhagic events.¹⁴⁹ This was a nationwide cohort study containing 182,678 individuals with a diagnosis of AF (ICD-10 code I489:A-F) who were treated as an inpatient or outpatient at Swedish hospitals between July 2005 and December 2008. Average follow-up was 1.5 years. This is a very large cohort, likely to be reflective of the general Swedish population with AF (although the sampling methods did exclude patients with 'silent AF' and patients managed only in primary care and open clinics). The other advantage of this study as a source of data for the model was that thromboembolic and bleeding events were reported for the same cohort, providing coherent estimates of these two related risks. The applicability of this data to the UK AF population is discussed below.

One-year incidence rates for thromboembolism (stroke/TIA/peripheral embolism) stratified by CHA₂DS₂VASc scores were reported for 90,490 patients not treated with warfarin (Table 25). Figures used in the model were adjusted for aspirin use to provide estimate rates for an untreated cohort. Figures for CHA₂DS₂VASc scores of 7 or more were pooled, as estimated rates were uncertain above this value - due to small numbers of events.

Similarly, incidence rates of major bleeds (intracranial and major extracranial) were reported by HAS-BLED scores. Rates used in the model were for 33,486 patients who were not on oral anticoagulation or aspirin at baseline. Event rates for HAS-BLED scores of 4 or more were pooled, due to small numbers of events.

Table 25. Rates of thromboembolism from the Swedish AF cohort study ¹⁴⁹

CHA ₂ DS ₂ -VASc score	n	Stroke/TIA/peripheral emboli Per 100 person years at risk (no warfarin, adjusted for aspirin ^o)
0	5,343	0.3
1	6,770	1.0
2	11,240	3.3
3	17,689	5.3
4	19,091	7.8
5	14,488	11.7
6	9,577	15.9
7-9	6,292	18.4
	90,490	7.0

* Estimates for patients not prescribed warfarin during follow-up, and adjusted for aspirin use

Table 26. Rates of bleeding from the Swedish AF cohort study ¹⁴⁹

HAS-BLED score	n*	Major bleeds Per 100 person years at risk (No oral anticoagulation or aspirin)
0	1,754	0.5
1	6,871	2.1
2	12,219	3.6
3	9,127	5.5
4-7	3,514	10.9
	33,486	2.1

* Estimated from distribution of HAS-BLED scores in the whole cohort (170,291) and number on no prophylaxis (33,486)

We assumed that the incidence of different types of thromboembolic event would be in the same proportions as observed in the Swedish AF cohort: 70% ischaemic strokes, 25% TIA, and 5% other embolisms. Similarly, the relative incidence of bleeds was also based on the observed rates in the Swedish AF cohort: 28% intracranial and 72% major extra-cranial. Minor bleeds were excluded from the model.

AF control and cardioversion

Data to populate the AF progression and control model (Figure 19) were drawn from two main sources. Euro Heart Survey data ¹⁶⁶ were used to derive estimates of the proportion of first episodes that are paroxysmal (42%), recurrence rates for paroxysmal AF (54% per year), rates of progression from paroxysmal to persistent AF (20% of recurrences), and the proportion of AF recurrences that are of acute onset (64%). The proportion of recurrences for patients with paroxysmal AF that are undocumented (68%) was taken from the Canadian Registry of Atrial Fibrillation study. ¹⁶⁷

Rates of recurrence for patients with persistent AF and progression from persistent to permanent AF were determined by effectiveness data for cardioversion (CG36 Chapter 5). These studies reported on the initial success rate of cardioversion (reversion to sinus rhythm within 24 hours), early recurrences (one day to two weeks) and late recurrences (beyond two weeks).

Table 27. Cardioversion success rates

	N	Sinus rhythm	Success rate	Standard error
Pharmacological cardioversion (first attempt)	158	123	0.78	0.0374
Electrical cardioversion (first attempt)	211	160	0.76	0.0338
PCV or ECV second attempt	37	23	0.62	0.1011

Estimated from studies reported in CG36: p35

Table 28. AF recurrence after cardioversion

	N	Recurrence	Recurrence rate	Standard error
Early recurrence (< one month)	171	53	0.31	0.0635
Late recurrence (> one month)	1023	502	0.49	0.0223

Estimated from studies reported in CG36: p35

Diagnostic accuracy

Sensitivity of the initial 12-lead ECG (78%) was based on the HTA trial by Hobbs *et al*¹²⁹, and it was assumed that 5-24% of patients with suspected paroxysmal AF would be diagnosed following a positive 24 hour or event recorder ECG.¹⁶⁸ Other probabilities within the diagnostic pathway – including the proportion of patients with a negative ECG referred for ambulatory assessment on the basis of suspicion of paroxysmal AF, the ratio of 24 hour to event recorder ambulatory ECGs, referral rates for TTE and TOE - were estimated by informal elicitation from experts.

Treatment effectiveness

Estimates of the effectiveness of antithrombotic medications (aspirin, warfarin, dabigatran and rivaroxaban) were taken from a network meta-analysis conducted by the BMJ Technology Assessment Group for the NICE TA on rivaroxaban.¹³⁶ This study estimated odds ratios for thromboembolic events (ischaemic stroke and systemic embolisms), bleeding (intra-cranial and major extra-cranial bleeds) and treatment withdrawals in comparison with warfarin. As inputs to the model, we estimated relative risks compared with placebo (see Table 29), using assumed control risks from the warfarin arm of the ROCKET AF trial.¹⁶⁹

Table 29. Treatment effects of anti-thrombotic drugs ¹³⁶

	Aspirin	Warfarin	Dabigatran	Rivaroxaban
Relative risk of thromboembolism *	0.45	0.30	0.23	0.25
Relative risk of bleeding *	1.12	1.76	1.63	1.81
Withdrawal rate (per person per year)	0.10	0.18	0.24	0.19

* compared with placebo

Treatment effects for the rhythm control drugs were taken from a published network meta-analysis, funded by Sanofi-Aventis to inform their submission to NICE for the appraisal of dronedarone.¹⁷⁰ For the model, we estimated relative risks for AF recurrence and treatment withdrawals (Table 30).

Table 30. Treatment effects of anti-arrhythmic drugs ¹⁷⁰

	Dronedarone	Amiodarone	Sotalol	Flecainide	Propafenone
Relative risk of AF recurrence	0.7938	0.4906	0.6948	0.6054	0.6576
Withdrawal rate (per person per year)	0.2847	0.2829	0.2453	0.3047	0.3047

* compared with placebo

The effectiveness of the rate control drugs was estimated by simulating an initial resting heart rate for each individual patient: sampled from a normal distribution with a mean of 109 and standard deviation of 31).¹⁷¹ Mean reductions in heart rate with rate control drugs were estimated from four randomised cross-over studies reported in CG36 (p59).¹⁷¹⁻¹⁷⁴ Data were pooled using a simple inverse-variance method of meta-analysis. Estimates of the standardised mean difference were obtained for first line treatment (beta-blocker (BB) or rate limiting calcium antagonist (RLCA)) compared with no treatment, and for second-line treatment (BB and digoxin, or RLCA and digoxin) compared with first line treatment. No data were found to estimate the effect of the third-line of treatment recommended in CG36 (amiodarone). We therefore assumed that this gives the same additional reduction in heart rate as second line treatment.

Table 31. Treatment effects of rate control drugs ¹⁷¹⁻¹⁷⁴

	N	Standardised mean difference	
		Mean	SE
First line (BB or RLCA) vs no treatment	34	0.87	0.20
Second line (BB+dig, RLCA+dig) vs. first line	55	0.70	0.15
Amiodarone vs. second line	<i>No data identified, assumed equal to second line</i>		

* compared with placebo

Case fatality rates

The proportion of patients admitted for acute AF dying with 30 days of admission was estimated at 2.56% (147 out of 5,735 cases in a study of all patients admitted with a diagnosis of AF in Scotland in 1996).¹⁷⁵ The case fatality for ischaemic stroke were estimated from a large population-based cohort study (OXVASC).¹⁷⁶ Fatality rates for bleeding were estimated from a pooled analysis of data from the SPORTIF III and V trials of ximelagatran versus warfarin for treatment of nonvalvular AF.¹⁷⁷ This estimated a case-fatality rate for major bleeding of 8.1% (very similar for the two study arms). The fatality rate amongst patients experiencing an intracranial bleed was much higher (10 out of 18 cases, 56%). A similar case-fatality rate was observed for haemorrhagic strokes in the OXVASC study: 8 out of 17 cases (47%).

Utilities

Health utility estimates were drawn from three sources. Firstly, baseline utility values for members of the public with no history of heart problems by 5-year age band were taken from the analysis of Health Survey for England data reported by Ara and Brazier¹⁷⁸, see Table 32.

Utilities were then adjusted for patient's AF status, using data from the RealiseAF study¹⁷⁹, which is an international observation cross-sectional study of patients with any history of AF in the previous year. Out of 9,665 patients evaluated, 26.5% were in sinus rhythm, 32.5% had an arrhythmia but with heart rate of ≤ 80 bpm and 41% had uncontrolled AF (neither in sinus rhythm for heart rate ≤ 80 bpm). EQ-5D scores were available for 9,644 of these patients. We used this data to estimate utility multipliers for AF status (see Table 33).

Table 32. General population utility by age (no other heart problems) ¹⁷⁸

Age group	N	Mean utility	95% CI	
<30	8040	0.9389	0.935	0.942
30 to ≤35	3592	0.9148	0.907	0.922
35 to ≤40	3992	0.9075	0.901	0.913
40 to ≤45	3703	0.8855	0.876	0.894
45 to ≤50	3243	0.8664	0.854	0.877
50 to ≤55	3089	0.8376	0.828	0.847
55 to ≤60	3173	0.8269	0.815	0.837
60 to ≤65	2580	0.8189	0.805	0.832
65 to ≤70	2784	0.8132	0.799	0.827
70 to ≤75	2276	0.7892	0.766	0.802
75 to ≤80	1709	0.7602	0.745	0.774
80 to ≤85	1072	0.7070	0.684	0.729
>85	572	0.6692	0.642	0.695

Table 33. Utility multipliers for AF health states, RealiseAF ¹⁷⁹

	N	Mean utility	Baseline	Multiplier
AF rhythm controlled (SR) vs no AF	2576	0.75	0.8132*	0.922
AF rate controlled (≤80bpm) vs AF (SR)	3123	0.72	0.7500	0.960
AF not in sinus rhythm vs AF (SR)	1014	0.67	0.7500	0.893
AF rate not controlled vs AF rate controlled	2931	0.67	0.7200	0.931

* Assumed average (Health Survey for England mean for 65-70 year olds)

Finally, we needed estimates of utility losses for the thromboembolic and bleeding events included in the model. These were obtained from the Medical Expenditure Panel Survey ¹⁸⁰, which collected EQ-5D scores for a nationally representative sample of 38,678 non-institutionalised adults between 2000 and 2002. This data was analysed using regression methods to estimate the marginal dis-utilities associated with 95 chronic conditions. Estimates of utility multipliers for adverse events included in the model are listed in Table 34.

Table 34. Utility multipliers for adverse events, MEPS¹⁸⁰

	N	Mean utility	Baseline	Multiplier
Ischemic stroke	38678	0.67	0.81	0.829
TIA *	38678	0.71	0.81	0.873
Systemic embolism*	38678	0.69	0.81	0.852
Haemorrhagic stroke	38678	0.67	0.81	0.829
Major bleeding *	38678	0.63	0.81	0.776

* TIA, other systemic embolism and bleeds assumed to be transient (acute period only)

Costs

Finally, estimates of costs for tests and treatments administered in the service pathway, drug costs, and costs for adverse events are shown in Table 35, Table 36 and Table 37 respectively.

Table 35. Costs for consultations and tests on the pathway

	2011 £	Source
GP visit	36.00	PSSRU ¹⁰⁵
Cardiologist first visit	175.00	DH Reference Costs ¹⁰⁴
Cardiologist follow up visit	122.00	DH Reference Costs ¹⁰⁴
Tertiary specialist visit	177.58	DH Reference Costs ¹⁰⁴
Emergency attendance	158.69	DH Reference Costs ¹⁰⁴
ECG	31.00	DH Reference Costs ¹⁰⁴
Event recorder ECG	45.00	DH Reference Costs ¹⁰⁴
24 hour monitor ECG	56.00	DH Reference Costs ¹⁰⁴
TTE or TOE	185.00	DH Reference Costs ¹⁰⁴
Cardioversion (PCV/ECV)	773.00	DH Reference Costs ¹⁰⁴

Table 36. Daily drug costs (NHS Net Price, BNF ¹⁰⁶)

		2011 £ (per day)
Antiplatelet	Aspirin	0.02
Anticoagulant	Warfarin	0.25
	Warfarin initiation (one off)	0.00
	Warfarin administration (per day)	0.66
	Dabigatran	2.52
	Rivaroxaban	3.03
	Heparin	6.96
Antiarrhythmic Class I	Class Ic	
	Flecainide	0.12
	Propafenone	0.35
	Flecainide (intravenous)	11.00
Antiarrhythmic Class II	Atenolol	0.05
	Labetalol	0.46
	Labetalol (intravenous)	2.94
	Esmolol (intravenous)	4.98
	Atenolol (intravenous)	3.60
	Esmolol+Digoxin	5.47
	Atenolol+Digoxin	0.54
	Labetalol+Digoxin	0.95
	Bisoprolol	0.03
	Metoprolol	0.13
Antiarrhythmic Class III Class III	Sotalol	0.18
	Amiodarone	0.18
	Amiodarone (intravenous)	7.13
	Dronedarone	2.25
Class IV (RLCA)	Diltiazem	0.16
	Verapamil	0.26
	Verapamil (intravenous)	1.64
Positive inotropic drug	Digoxin	0.49

Table 37. Cost of adverse events included in model

		2011 £	Source
Ischaemic Stroke	Event cost (one off)	14,426	<i>Luengo-Fernandez</i> ¹⁷⁶
	Ongoing cost (per day)	23.48	<i>Luengo-Fernandez</i> ¹⁷⁶
	Acute period (days)	90	<i>Luengo-Fernandez</i> ¹⁷⁶
TIA	Event cost (one off)	402	<i>DH Reference Costs</i> ¹⁰⁴
	Ongoing cost (per day)	0	<i>Assumption</i>
	Acute period (days)	1	<i>Mean length of stay</i>
Other thromboembolism	Event cost (one off)	1,705	<i>DH Reference Costs</i> ¹⁰⁴
	Ongoing cost (per day)	0	<i>Assumption</i>
	Acute period (days)	6	<i>Mean length of stay</i>
Haemorrhagic stroke	Event cost (one off)	16,228	<i>Luengo-Fernandez</i> ¹⁷⁶
	Ongoing cost (per day)	15.75	<i>Luengo-Fernandez</i> ¹⁷⁶
	Acute period (days)	90	<i>Luengo-Fernandez</i> ¹⁷⁶
Other major bleed	Event cost (one off)	725	<i>DH Reference Costs</i> ¹⁰⁴
	Ongoing cost (per day)	0	<i>Assumption</i>
	Acute period (days)	5	<i>Mean length of stay</i>

Verification and validation

The model was coded using SIMUL8 software by one of the authors (JE). The data spreadsheet was prepared by another author (MTB), who worked closely with JE to ensure that the data interface worked correctly. Data entry was checked by another author (JL), by comparison of data from the original papers with the numbers in the spreadsheet. The modelling team (MTB, JE and JL) met regularly to discuss and resolve problems arising. At several stages in development, the model was discussed at meetings of the wider MAPGuide team, and issues about model structure, coding and data sources were considered.

Various steps were taken during model development to avoid potential errors. This included double coding of some of the more complicated formulae in Excel and SIMUL8, to check that they were being applied correctly. This included: the method for sampling time to event and for dealing with competing risks; the formulae used to calculate individuals' risk of coronary heart disease, diabetes and hypertension (the Framingham formulae); and the formula for continuous discounting of costs and QALYs. A patient diary was also created to collect information about the events and timelines for individual patients running through the model, and to present this in the form of individual 'case histories'. These diaries were used throughout

development, and towards the end of this process case histories were checked for 500 patients by members of the modelling team, and any apparent inconsistencies or errors were identified, discussed and if necessary investigated.

Verification of the coding was commissioned to an external modeller based at the University of Sheffield. This comprised three tasks:

- (a) Checking that the SIMUL8 logic correctly reflected the AF pathway;
- (b) Checking the coding of costs, QALYs and discounting calculations; and
- (c) Checking the model logic via the patient diaries.

Finally, the model outputs were compared against the model inputs to ensure that the rates of thromboembolic events and bleeds observed for a modelled cohort matched the input data ('internal validation'). The results of the model were also analysed to verify that the relative risks for different antithrombotic drugs correctly reflected the data built into the model.

The following list summarises the set of assumptions that were agreed during model development.

Summary of model simplifications

- 1) Patients' individual blood pressure and cholesterol levels remain constant throughout.
- 2) Patients have hypertension if they are taking hypertensive drugs at the start—regardless of their blood pressure
- 3) As the ONS data only cover ages up to 100, patients aged over 99 are not considered eligible for this model.
- 4) Event recorder and 24-hour ambulatory ECG tests have a specificity of 1 (i.e. no false positives as it is unlikely that an arrhythmia will be picked up if it does not exist), but have varying sensitivities to allow for false negatives.
- 5) By definition, no-one can be classified as in permanent AF state on initial diagnosis as they will not have had the opportunity to try cardioversion at this point
- 6) The contraindicated label encapsulates any reason that the patient cannot take the particular drug, so includes reactions, treatment failure, and ineligibility due to personal characteristics.

- 7) Patients continue with the final line of treatment after referral to the super-specialist, but continue in the model accumulating costs and QALYS until their death. This section is outside the scope of the guideline, but it was felt that the costs and QALYS should be included in the analysis
- 8) No time passes while patients pass through the classification, stroke risk classification, acute onset, cardioversion, rate control and rhythm control sections of the model. The reason for this assumption is that we are not currently modelling resources explicitly, and as costs are accumulated on a daily basis, times of less than 1 day will not change the results.
- 9) We are not considering the costs, impact and effects of medications for co-morbidities within the model, so only the medications and anti-coagulants that are prescribed as part of the treatment for AF are considered.

Modelling pathway changes

The approach to evaluating the selected topics by the AF model are outlined in Table 38, and discussed in more detail below.

Table 38. Topics for atrial fibrillation model

Topic	Description	Place in pathway	Options to be evaluated	Additional data and assumptions
A	Prophylaxis for the prevention of post-operative AF	Pre-diagnosis	None	Model does not cover post-operative AF.
B	Anti-arrhythmic drugs as pharmacological cardioversion (PCV) for people with atrial fibrillation	PCV for AF onset <48 hours, no structural heart disease (C11)	B0) Class 1c drug B1) Amiodarone	Success of cardioversion for class 1c drugs and amiodarone at 8 and 24 hours ¹⁸¹ . Assume one extra bed day if sinus rhythm is not restored within 8 hours, at an additional cost of £338 ¹⁰⁴ . Complications of PCV procedure added ¹⁸² , assumes 1.78 extra bed days per complication ¹⁰⁴ .
C	Rhythm versus rate control for persistent AF; subgroups including those with hypertension, previous MI and congestive heart failure	Classification (CL1)	C0) CG36 criteria C1) Rhythm for all persistent C2) Rate for all persistent	Intended as simple illustration of how this question could be addressed. A fuller analysis would involve addition of cardiac and other serious adverse effects for both rate and rhythm control drugs ^{170;183;184} , and costs and quality of life impacts for these adverse effects ¹⁴⁰⁻¹⁴² . A key uncertainty is whether sinus rhythm has an independent effect on thromboembolisms ¹⁸⁵ . Analysis for subgroups would require relative risks for various modeled outcomes and adverse effects.
D	Treatment for maintaining Sinus Rhythm in people with AF after cardioversion.	Rhythm control	Not run due to time constraints, but feasible.	Comparison of alternative drugs for rhythm control would require addition of adverse effects as mentioned above.
E	Alternative risk factor based scoring systems to estimate stroke and embolism risk	Stroke risk	E0) Warfarin for all E1) Warfarin if CHA2DS ₂ -VASc ≥ 1 E2) Warfarin if CHA2DS ₂ -VASc ≥ 2 E3) Warfarin if CHA2DS ₂ -VASc ≥ 3 E4) Warfarin if CHA2DS ₂ -VASc ≥ 4 E5) Warfarin if CHA2DS ₂ -VASc ≥ 5 E6) Warfarin if CHA2DS ₂ -VASc ≥ 6 E7) Warfarin if CHA2DS ₂ -VASc ≥ 7 E8) Warfarin if CHADS ₂ ≥ 1	No additional data needed. To simplify the interpretation of results, these analyses compare only warfarin and aspirin, but other drugs could be added. All analyses assume that patients who have a bleed on warfarin will switch to aspirin, no contraindications to aspirin were included, although these could be added.

Topic	Description	Place in pathway	Options to be evaluated	Additional data and assumptions
			E9) Warfarin if CHADS ₂ ≥ 2 E10) Warfarin if CHADS ₂ ≥ 3 E11) Warfarin if CHADS ₂ ≥ 4 E12) Warfarin if CHADS ₂ ≥ 5 E13) Aspirin for all	
F	Stratification tools to assess bleeding risk before prescription of antithrombotic medication	Stroke risk	F0) Warfarin for all F1) Warfarin only if HAS-BLED < 4 F2) Warfarin only if HAS-BLED < 3 F3) Warfarin only if HAS-BLED < 2 F4) Warfarin only if HAS-BLED < 1 F5) Aspirin for all	As above.
G	Apixaban, rivaroxaban or Dabigatran etexilate versus warfarin as for patients at moderate or high risk of stroke or systemic embolism	Stroke risk	G0) Warfarin/dabigatran/ rivaroxaban/ aspirin (CG36 + TA249 and TA256) G1) Warfarin/aspirin (CG36)	As an illustration, we compare results using the original CG36 recommendations, and with the addition of recommendations from TA249 and TA256. No additional data is needed.
H	Cather ablation for paroxysmal and persistent AF patient	Rhythm control	Not run due to time constraints, but feasible.	Model could be adapted by adding percentage of patients referred who have procedure, success and recurrence rates (similar to cardioversion), costs and adverse events of procedure.

Topic A: Prophylaxis for prevention of postoperative AF

Onset of AF following cardiothoracic surgery is a common problem. This is sometimes transient, but AF can persist and if so is associated with potentially serious effects: including haemodynamic instability, ischaemia, heart failure and stroke and thromboembolism ¹²⁷. CG36 recommends the use of prophylaxis to prevent postoperative AF (amiodarone, beta-blockers, sotalol or rate-limiting calcium antagonist) and management with an initial rhythm-control strategy (Chapter 10). The 2011 review ¹⁸⁶ highlighted new evidence related to the choice of prophylactic treatment (including statins and corticosteroids as well as anti-arrhythmic drugs), the timing of prophylaxis (pre, intra or post-operative), and subsequent treatment.

However, the modelling team did not include post-operative AF in the base case model because the clinical and demographic characteristics of patients undergoing cardiothoracic surgery and their risks of adverse events differ from those of the general AF population. Although the base case model could be adapted to reflect post-operative AF, this would require further consultation with experts, identification of suitable evidence sources for both baseline risks and treatment effects and possibly re-structuring of the model. Consequently this topic was not evaluated.

Topic B: Drugs for pharmacological cardioversion

Background to topic

CG36 ¹²⁷ made recommendations for the cardioversion of patients in AF, including patients presenting as an emergency with haemodynamic instability (Chapter 7) and stable patients for whom a rhythm control strategy is being pursued (Chapter 5). For the latter more common situation, the evidence suggested that Electrical Cardioversion (ECV) and Pharmacological Cardioversion (PCV) were of comparable efficacy. Based on clinical practice and opinion, the guideline group recommended ECV for prolonged AF (episode lasting for 48 hours or longer), and either PCV or ECV for AF of more recent onset. Evidence on the relative effectiveness of drugs for PCV suggested that although the Class 1c anti-arrhythmic drugs (flecainide and propafenone) are more effective than amiodarone in the short term, they achieve a similar rate of conversion to sinus rhythm by 24 hours. The guideline group noted concerns over the safety of Class 1c drugs in patients with structural

heart disease (coronary artery disease or left ventricular dysfunction). They therefore recommended amiodarone for PCV in patients with structural heart disease, and a Class 1c drug for other patients. The group also reviewed evidence relating to adjuncts to ECV, including concomitant use of antiarrhythmic drugs (AADs) to increase the chance of success and reduce recurrence, and anticoagulation and transoesophageal echocardiography (TOE) to reduce the risk of stroke and thromboembolism associated with the ECV procedure.

The 2011 review¹⁸⁶ identified new evidence relating to the choice of methods for cardioversion as part of a rhythm control strategy. This included studies comparing drugs for PCV, ECV with concomitant AAD treatment, and different ECV techniques. Four RCTs were found relating to a new Class III agent vernakalant, which is licensed for cardioversion of AF episodes lasting for 7 days or less. This drug was referred to NICE as a Single Technology Appraisal, but the process was suspended due to information about the timing of launch in the UK. Vernakalant is not yet available in the UK and does not have an NHS price, so evaluation is not currently possible. None of the new evidence relating to other drugs for PCV identified in the review document provides a suitable basis for economic evaluation, since they have not demonstrated superiority over current recommendations. Similarly, it is difficult to identify a clear evidence base to test potential changes in CG36 recommendations on concomitant AAD treatment with ECV, or strategies to prevent thromboembolism associated with ECV.

Approach to economic evaluation

We therefore decided to use existing evidence included in CG36 to illustrate how the model could be adapted for evaluation of different methods of cardioversion. As an example, we focus on the choice of PCV drugs in patients with AF onset of less than 48 hours, without haemodynamic instability or structural heart disease (recommendation R8 p40 CG36¹²⁷): comparing the use of Class 1c drugs (as is currently recommended) with amiodarone.

Key factors likely to drive the cost-effectiveness of different methods of cardioversion are: the frequency and speed of conversion to sinus rhythm; the frequency and severity of adverse events; the risks of future recurrence; and the cost of the procedure, consumables and associated hospital stay. The efficacy and speed of cardioversion is clearly important for patients; as faster resolution of symptoms will alleviate discomfort, distress and anxiety. The short duration of this benefit and the lack of quality of life data make it difficult to

capture in a QALY metric. However, speed of cardioversion is likely to impact on the use of NHS resources and costs, which should be easier to estimate. Although we have not identified any direct evidence that more rapid cardioversion translates to a shorter hospital stay, this is a reasonable inference. We therefore modelled this effect by applying an additional cost for patients not converted to sinus rhythm by a given time, to reflect the likelihood of an extra bed day.

In addition to the efficacy and speed of cardioversion, the risk of adverse events is a major factor that influences clinicians' choice of drugs for PCV. In particular with antiarrhythmic therapy there is the risk of proarrhythmia, where the treatment itself can precipitate the onset of a new arrhythmia, including bradycardia, tachycardia or prolongation of the QRS or QT interval. As with any drugs, AADs may be associated with a range of other adverse events, including headache, nausea, dizziness and ocular disturbances. Though most adverse events have no long term consequences, they can be unpleasant and are potentially harmful. As is often the case, modelling of adverse events is difficult because of the wide variety of types and severity of events associated with the rhythm control drugs. Data on adverse events is also sparse and difficult to collate due to variations in how it is reported. Nevertheless, it is important that they can be included in evaluations of AADs, both in the context of short-term use for PCV (as in this topic) and in ongoing treatment (as in the following two topics). In the illustrative analysis presented below, we included an additional cost to reflect a longer length of stay for patients experiencing a complication during PCV. We did not include any QALY loss for complications, as none of the complications observed in the identified trials had any lasting effect.

We are not aware of any evidence to suggest that the choice of drug for PCV would influence the risk of recurrence in patients successfully converted to sinus rhythm, and so did not include this in our illustrative evaluation on this question. CG36 did present evidence that concomitant use of AADs alongside ECV can reduce the rate of relapse to AF (Tables 5.5 and 5.6 p42 CG36). It would be easy to incorporate any such impacts by applying appropriate relative risks to the early and late recurrence rates following successful cardioversion.

Sources of data

The guideline recommendations on the choice of PCV drug were informed by a meta-analysis¹⁸¹ comparing amiodarone, class 1c drugs (propafenone or flecainide) and placebo. This review concluded that the Class 1c drugs were more effective than amiodarone at achieving sinus rhythm by 8 hours (63% versus 42%, $p < 0.001$), but that there was no significant difference by 24 hours (71% versus 66%, $p = 0.50$). We assumed that after 8 hours, one extra bed day would be incurred at a cost of £338 (the excess bed day cost for an elective inpatient stay for HRG EB071¹⁰⁴). For comparison, a randomised trial of cardioversion¹⁸⁷ reported a mean length of stay of one day ($sd = 2$) for 72 patients randomised to an initial PCV strategy, and 2 days ($sd = 2$) for 67 randomised to initial ECV. The review by Chevalier and colleagues¹⁸¹ also considered reports of adverse events, but found insufficient data to quantify the risks. Another review¹⁸² conducted by the manufacturer of the new drug vernakalant, collated safety data from 22 trials of a range of PCV drugs, including amiodarone, class 1c drugs, as well as vernakalant. They reported an overall adverse event rate within 2 hours of 40% (188 of 472 patients), and a serious adverse event rate of 4% (25 of 637). Four deaths were reported during this initial period and a further eight after 24 hours, although it was reported that none of these deaths were related to treatment. There were insufficient data to estimate relative adverse event rates for different drugs. For the PCV evaluation, we assumed that 4% of patients undergoing PCV would experience a serious complication, incurring an additional 1.78 days in hospital at a cost of £338 per day (total cost £602 per complication). This estimate is based on the difference in average length of stay for elective inpatient episodes for arrhythmia or conduction disorders (EB07) for patients with/without complications, and the cost of an excess bed day for these patients¹⁰⁴. For simplicity we assumed that other more minor side effects would not have any significant health or resource impact. A relative risk parameter was added to the model to allow sensitivity analysis of the impact of a difference in the percentage of patients experiencing complications with different drugs.

We ran the model twice, once assuming that patients without structural heart disease undergoing PCV would receive an intravenous Class 1c drug, and once assuming that they would receive intravenous amiodarone. The difference in net benefit between these

strategies illustrates the potential for gain from identifying faster acting PCV drugs for use in this context.

Topic C: Rhythm versus rate control for persistent AF

Background to topic

The choice of treatment strategy for patients with persistent AF is controversial. There is no clear evidence that patients with persistent AF benefit from attempts at regaining and maintaining sinus rhythm through cardioversion and use of AADs, or whether they would achieve better outcomes by moving straight to a rate control strategy. CG36 reported that no study had demonstrated rhythm control to be superior to rate control (or vice versa) for the outcomes of mortality or quality of life, although the GDG concluded that there was 'generally consistent' evidence that rates of adverse events and hospital admissions were higher with a rhythm control strategy. The GDG also considered subgroup analyses of the AFFIRM trial¹⁸⁴, which identified some factors associated with a lower risk of death with rate control. More recent Cochrane reviews^{188;189} have reached similar overall conclusions.

The GDG highlighted difficulties in interpreting the evidence base in this field because of confounding with anti-thrombotic therapy. Trials comparing rate and rhythm control suffered from an imbalance between the arms in the proportion of patients treated with an oral anticoagulant (OAC), as patients allocated to rhythm control were often withdrawn from OAC treatment if they remained in sinus rhythm. The use of composite outcome measures makes it difficult to tease out the effects of rate/rhythm control from the effects of anti-thrombotic therapy. The primary outcome for most rate versus rhythm trials has been all-cause mortality, which includes deaths from thrombotic, haemorrhagic and arrhythmic events. Rates of hospitalisation are also confounded, since they often included admissions for cardioversion, as well as for treatment of adverse events and treatment-related side effects.

Given this equivocal evidence, CG36 recommended that the choice of strategy should be tailored for individual patients, and suggested criteria to guide this choice. The criteria for initial rate control were age (over 65), coronary artery disease, absence of congestive heart failure, and suitability for cardioversion and AADs. Additional criteria for rhythm control

were symptomatic AF, first presentation of lone AF, and AF secondary to a treated/corrected precipitant.

The review of CG36¹⁸⁶ identified some evidence that could potentially be used to revise criteria for rate versus rhythm control in defined patient groups:

- An analysis of data from the RACE trial¹⁹⁰ reported that hypertensive patients randomised to rhythm control were at greater risk of cardiovascular morbidity and mortality than those randomised to rate control. This difference was not seen for non-hypertensive patients.
- The AF-CHF trial¹⁹¹ reported that patients with heart failure randomised to rhythm control had a similar risk of death from cardiovascular causes compared with those randomised to rate control.
- The VALIANT study¹⁹² compared rate or rhythm control strategies in patients following a myocardial infarction. This found an excess mortality associated with anti-arrhythmic drugs over the first 45 days, but no increased mortality after this initial period.

Approach to economic evaluation

It is not possible to use the direct rate versus rhythm evidence in the MAPGuide model, because of the problems of confounding with OAC treatment and composite outcomes. The model design separates non-AF related mortality which is defined at model entry and independent of interventions and events in the pathway, and AF-related mortality which is defined as a case-fatality rate, consequent on the occurrence of certain events (acute arrhythmia, thromboembolism or bleeding), as shown in. The model is therefore incompatible with data on all-cause mortality effects. Similar problems apply to data on rates of hospitalisation when it is not possible to separate the reasons for admissions. This problem applies to the subgroup analyses highlighted in the guideline review, as well as to the overall comparison of rate versus rhythm control.

There are two mechanisms through which rhythm and rate control interventions impact on health outcomes in the model: through acute arrhythmic episodes that can be fatal, and through the reduced quality of life associated with uncontrolled AF (lack of sinus rhythm or heart rate above 80bpm). It is straightforward to use these existing modelled mechanisms

to compare the overall cost and QALY impact of directing all patients with persistent AF to rhythm control or to rate control. The results of this analysis are reported below. However, this analysis suffers from some important limitations, particularly the rather limited incorporation of adverse drug effects within the current base case model. These issues are addressed in the Discussion section below.

Topic D: Antiarrhythmic drugs

CG36 recommended an escalating sequence of drugs, starting treatment with a standard beta-blocker. In patients without structural heart disease for whom a beta-blocker is ineffective, contraindicated or not tolerated, the GDG recommended use of a class 1c agent or sotalol. For patients with structural heart disease, or in patients for whom other treatment options have failed, amiodarone was recommended. This sequence was largely based on concerns over adverse effects, rather than on efficacy or cost. Evidence suggests that amiodarone is the most effective drug for maintaining sinus rhythm, but it is associated with some potentially serious adverse effects, including pulmonary, hepatic, ophthalmic and thyroid toxicity. The guideline review¹⁸⁶ highlighted recent evidence relating to new and existing antiarrhythmic agents. New treatments included dronedarone, which had been the subject of a NICE technology appraisal¹⁴¹, ACE inhibitors and ARBs, and rosuvastatin.

Similar issues arise for the evaluation of particular antiarrhythmic drugs as for the comparison of rate versus rhythm control strategies discussed above. In particular, adverse effects are expected to have an important influence on the choice of antiarrhythmic drug. However, although the current version of the base case model includes treatment withdrawals, it does not account for any significant or lasting costs or health impacts for adverse effects. We therefore consider that the base case version of the model is not a suitable platform for evaluation of this topic. Adaptation of the model to incorporate these effects is possible (see Discussion below), but there was insufficient time in this project to make the necessary changes. Consequently, we did not attempt to use the model to evaluate this topic.

Topic E: Risk factor scoring systems for stroke and embolism risk

CG36¹²⁷ includes a stroke risk stratification algorithm that prioritises patients for oral anticoagulation. It groups patients into high, medium and low risk groups on the basis of

their age and history of ischaemic stroke/TIA or other thromboembolic event, hypertension, diabetes, vascular disease, valve disease, heart failure or left ventricular dysfunction. This was based on a review of evidence on individual risk factors, and of existing risk stratification algorithms.

The guideline review¹⁸⁶ identified various new studies defining and testing different stroke risk stratification schemes for patients with AF. The European Society of Cardiology (ESC) published its guideline on atrial fibrillation in 2010¹⁹³. This recommended the CHADS₂ scheme as a simple initial method for assessing stroke risk, with patients scoring 2 or higher recommended for oral anticoagulation. For patients with a CHADS₂ score of 0 or 1, they recommend more detailed assessment with the CHA₂DS₂VASc scoring system.

The base case model incorporated CHA₂DS₂VASc as the means of stratifying stroke and thromboembolism risk, and applied the NICE criteria as the means of identifying patients suitable for oral anticoagulation. The model also included calculation of CHADS₂ scores in order that they could be used for comparison. It is therefore straightforward to compare these different scoring systems, and to compare different thresholds for prescription of oral anticoagulants.

Topic F: Stratification tools to assess bleeding risk

The review group also considered systems for assessing patients' risk of bleeding, which could be used to identify patients who would not be suitable for oral anticoagulation. CG36 listed criteria for assessing the risk of bleeding including age, use of antiplatelet drugs or NSAIDs, multiple other drugs treatments, uncontrolled hypertension, history of bleeding or poorly controlled anticoagulation therapy. However, recommendations about how these factors should be combined or evaluated were unclear. Since publication of the NICE guideline, more formal systems for assessing individuals' risk of bleeding have been developed, including the HAS-BLED scoring system, which was used in the model.

As with stroke risk assessment, it is straightforward to use the model to evaluate the application of HAS-BLED thresholds to limit the use of oral anticoagulation in patients at high risk of a bleed.

Topic G: Anticoagulant drugs

Since publication of CG36, two new oral anticoagulants have been recommended by NICE Technology Appraisals: dabigatran (TA249) issued in March 2012¹³⁷, and rivaroxaban (TA256) issued in May 2012¹³⁸. Both drugs were recommended with certain restrictions. Rivaroxaban was recommended as a treatment option for atrial fibrillation without underlying heart valve disease and at least one of the following additional risk factors:

- congestive heart failure
- high blood pressure
- age 75 or older
- diabetes
- or history of stroke or transient ischaemic attack.

Criteria for access to dabigatran were similar: non-valvular atrial fibrillation and at least one of the following:

- stroke, TIA or embolism in the past
- heart failure of class 2 or above
- age 75 or older
- age 65 or older with diabetes, coronary artery disease or high blood pressure.

The model already includes the above criteria, so the addition of dabigatran and rivaroxaban in line with NICE TA recommendations to the CG36 treatment pathway is straightforward, and does not require any additional data. The model makes certain assumptions about the proportion of patients receiving the different antithrombotic drugs when they are eligible for more than one drug (as illustrated in Table 23. In particular, it assumes equal use of aspirin and warfarin when both are recommended (e.g. for patients at medium risk); it assumes equal use of warfarin and of either dabigatran or rivaroxaban when all three are recommended; and it assumes equal use of rivaroxaban and dabigatran when both are appropriate. These criteria are easily changed, and a wide range of other strategies for the use of oral anticoagulants could be tested within the current model structure.

We have not attempted to model apixaban for use as an oral anticoagulant for AF, as this is scheduled for a NICE technology appraisal, not due to report until April 2013.

Topic H: Catheter ablation for paroxysmal and persistent AF

CG 36 includes recommendations for the referral of patients for consideration for various interventional procedures after failure of medical therapies for rate or rhythm control. This included referral for pulmonary vein isolation, which includes catheter-based procedures. However, the guideline did not review evidence for the effectiveness or cost-effectiveness of any specialist invasive procedures, or make recommendations for their use *per se*.

The guideline review identified new evidence relating to a range of ablation techniques, indicating an increased interest in this approach. Ablation is potentially curative of AF, although it is associated with a range of complications, including cardiac tamponade, pulmonary stenosis, thrombotic and haemorrhagic events. The HTA systematic review and cost-effectiveness analysis¹⁴³ concluded that radio frequency catheter ablation is more effective than antiarrhythmic drug therapy in patients with refractory paroxysmal AF, over a period of up to 12 months. However, evidence beyond 12 months is lacking, as is evidence relating to use of the technique in patients with persistent or permanent AF. The HTA economic analysis found that treatment would be cost-effective if quality of life improvements are maintained over the remaining lifetime of the patient, but that cost-effectiveness is unclear if the benefits are only maintained for five years. Other uncertainties relate to the effect of ablation on the risk of thromboembolism.

The costs and outcomes of ablation were not included in the base case model – as this was outside the original scope of CG36. Adaptation of the model to evaluate this procedure would be relatively straightforward, and would involve the addition of a pathway for selection of individuals for treatment, inclusion of the costs of treatment and follow up, and the costs and QALY impacts of complications, and adaptation of the AF recurrence rate calculations. As noted above, there might be some difficulty in identifying evidence of the longer-term impact of ablation on AF recurrence and thromboembolism risk.

Results

Base case scenario: deterministic results

Results for 10,000 patients sampled from the THIN AF cohort are shown in Table 39 and in Figure 22 to Figure 26. At diagnosis, this sample had an average age of 74, a mean CHA₂DS₂-VASc score of 3.3 and a mean HAS-BLED score of 2.3. The health outcomes and costs shown

are the results of one run of the base case simulation model, following the CG36 recommendations and using point estimates for all model parameters. These illustrate the magnitude and range of key outputs under the base case scenario. Mean predicted survival was 11 years, with a range from 0 to 66 years. Adjusting for utility during this time (mean 0.64), resulted in an estimated mean lifetime accumulation of around 7 QALYs (undiscounted). Over their lifetime, the simulated patients experienced a mean 0.31 thromboembolic events and 0.79 haemorrhagic events. Their AF treatment included a mean of 0.31 attempts at cardioversion, 0.16 acute admissions, and 2.68 tertiary consultations. The model predicted wide variations in costs, with the overall lifetime cost of AF-related care amounting to a mean of £28,230 per patient, rising to a maximum of £678,741. The cost of medication made up around 21% of this total cost. On an annual basis, the mean cost of medications was £538 per patient and the mean cost of other healthcare was £1,856 per patient (including costs for AF admissions and consultations, and treatment and care following AF-related adverse events).

To inform the decision over the number of patients to include per iteration in the probabilistic sensitivity analysis, we estimated the cumulative means of key output parameters for increasing numbers of patients. Figure 27 shows how the volatility of the estimated net benefit per patient decreases as the number of patients is increased to 10,000. The estimate is reasonably stable at 1,000 patients, and there is very little change in the estimate above 2,000 patients. For the analyses below, we used 1,000 patients per probabilistic iteration to limit the runtime of the model, as we wanted to compare a large number of scenarios.

Table 39. Results for 10,000 patients: base case model, deterministic

		<i>Mean</i>	<i>sd</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Age at Arrival		74	12	31	75	99
Initial CHA ₂ DS ₂ -VAsC score		3.3	1.7	0	3	9
Initial HAS-BLED score		2.3	1.2	0	2	9
Life Years	<i>Lifetime total</i>	11.2	9.2	0.0	9.0	65.8
QALYS	<i>Lifetime total</i>	7.15	5.88	0.00	5.69	38.02
	<i>Mean per year</i>	0.64	0.06	0.35	0.64	0.84
Thromboembolic events	<i>Lifetime total</i>	0.31	0.68	0	0	7
Haemorrhagic events	<i>Lifetime total</i>	0.79	1.06	0	0	8
Cardioversions	<i>Lifetime total</i>	0.36	0.80	0	0	12
Acute AF episodes	<i>Lifetime total</i>	0.16	0.44	0	0	7
Tertiary reviews	<i>Lifetime total</i>	2.68	6.59	0	0	73
Medication costs (£)	<i>Lifetime total</i>	5,989	6,762	0	3,817	70,293
	<i>Mean per year</i>	538	400	0	411	1,931
Other healthcare costs (£)	<i>Lifetime total</i>	22,240	49,291	0	3,204	671,148
	<i>Mean per year</i>	1,856	3,431	0	360	94,386
Total cost (£)	<i>Lifetime total</i>	28,230	50,823	0	10,709	678,741
	<i>Mean per year</i>	2,395	3,421	0	1,143	95,154

Figure 22. Age at diagnosis: 10,000 patients from THIN AF cohort

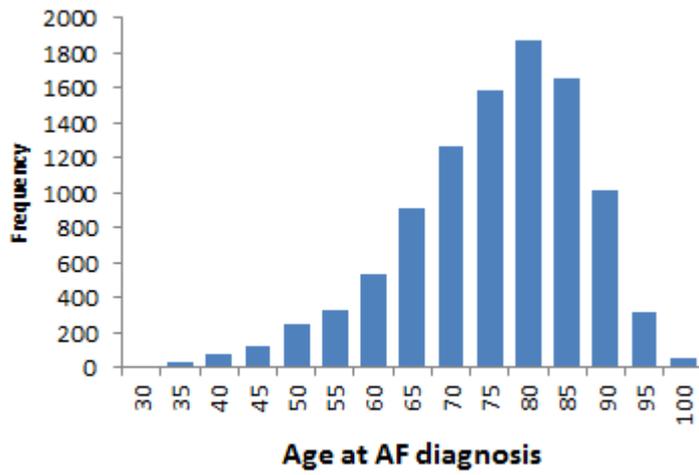


Figure 23. Simulated life years: 10,000 patients from THIN AF cohort

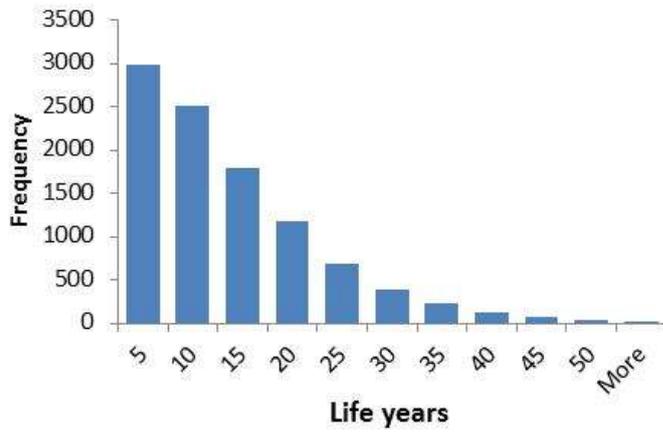


Figure 24. Simulated QALYs: 10,000 patients from THIN AF cohort

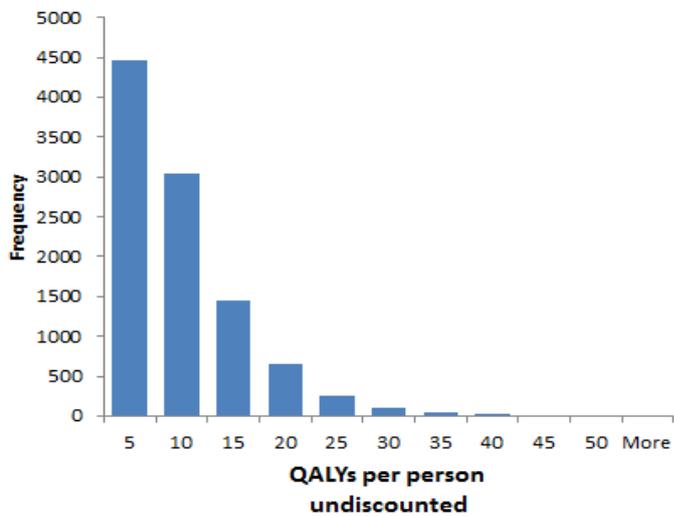


Figure 25. Simulated medication costs: 10,000 patients from THIN AF cohort

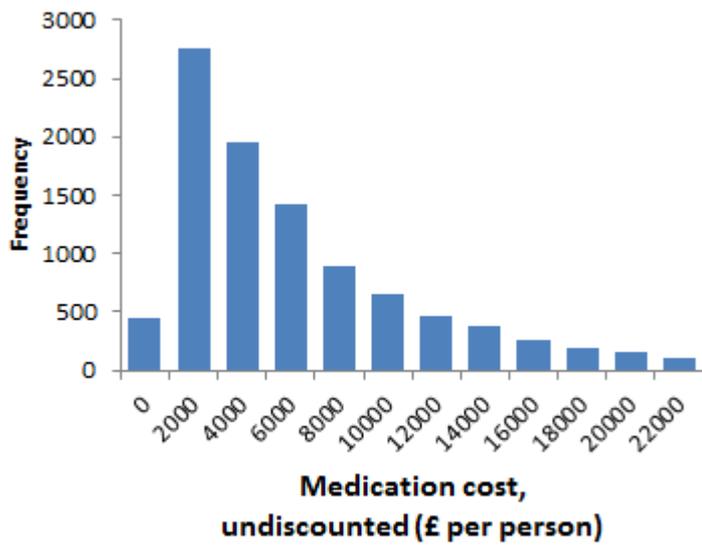


Figure 26. Simulated healthcare costs: 10,000 patients from THIN AF cohort

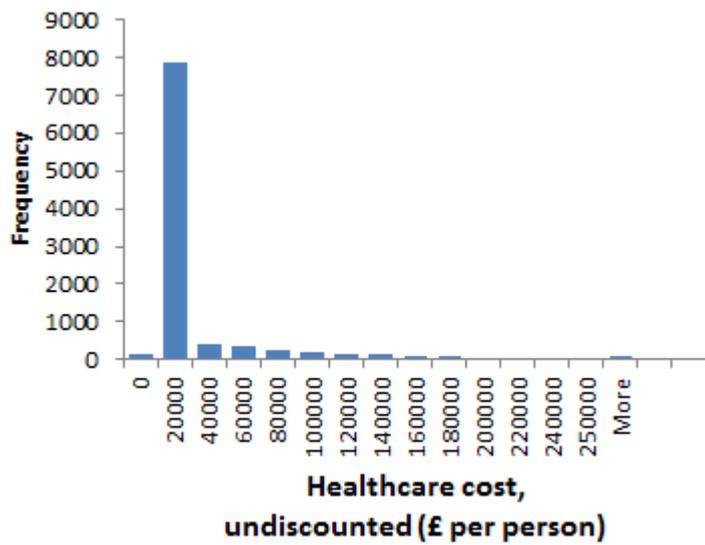
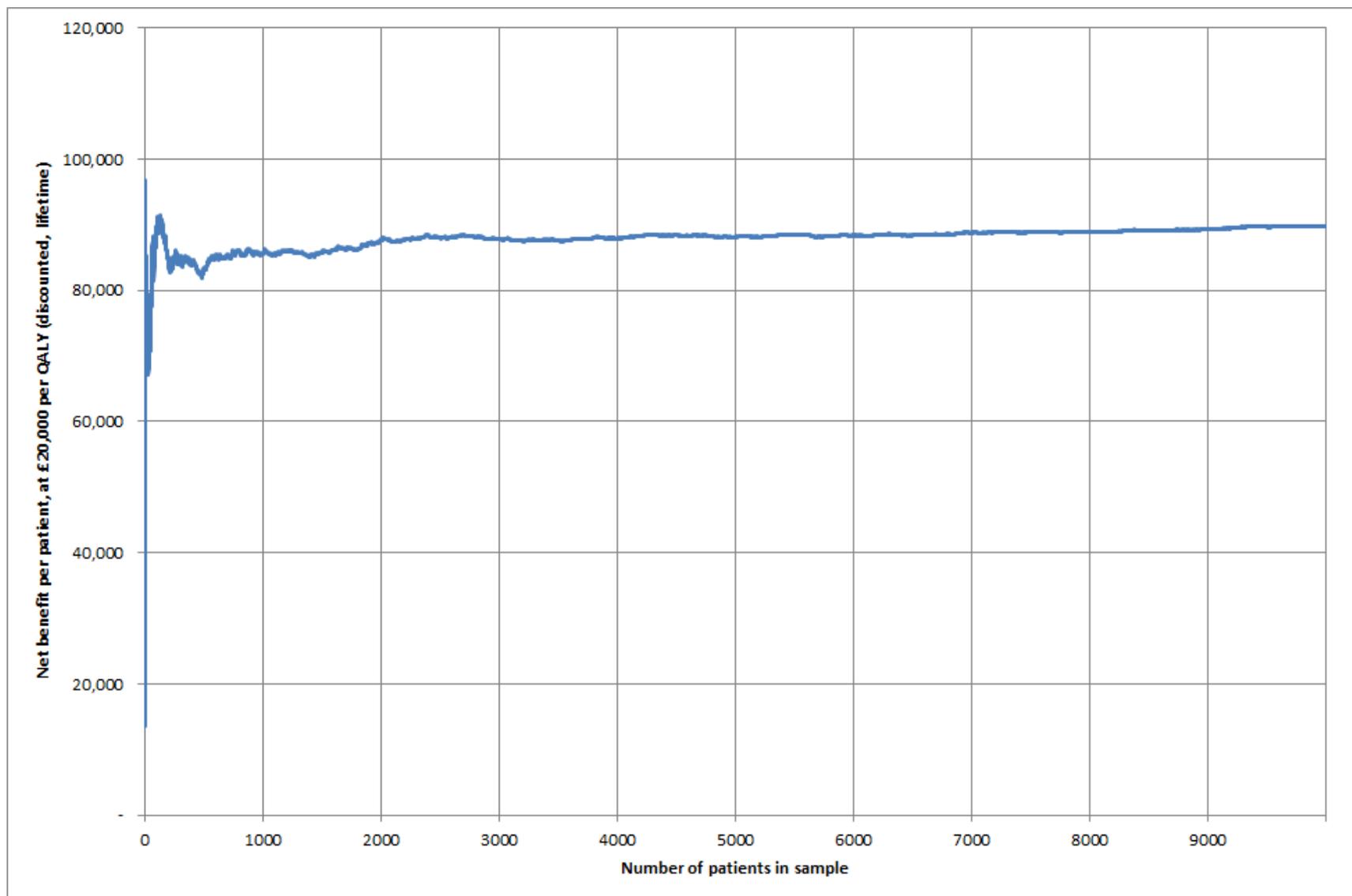


Figure 27. Cumulative net benefit per patient with increasing sample size



Base case scenario: probabilistic results

Table 40 compares the deterministic and probabilistic results for the base case scenario. With 500 probabilistic iterations and 1,000 patients per iteration, the results of the probabilistic and deterministic analyses were quite similar. The probabilistic sensitivity analysis produced an estimate of 5.22 QALYs per patient (discounted), compared with 5.20 QALYs from the deterministic analysis. There was rather more of a difference in estimated costs (£21,048 from the probabilistic analysis, compared with £19,494 from the deterministic analysis). This is not unexpected, as a large proportion of costs in the model related to treatment of relatively but expensive events (mainly ischaemic and haemorrhagic strokes). Overall net benefits were similar when estimated from the probabilistic and deterministic analyses: £83,441 versus £84,497 respectively (assuming a cost-effectiveness threshold of £20,000 per QALY).

The number of probabilistic iterations used for the analyses presented below was based on observation of the volatility of key output estimates with increasing numbers of iterations. Figure 28 shows how the expected net benefit per patient changed as the number of probabilistic iterations was increased from 0 to 500 for three illustrative strategies for antithrombotic therapy. It can be seen that the expected net benefit per patient is lowest with no antithrombotic treatment; the base case NICE strategy gives the next highest expected net benefit; and prescription of aspirin for all patients gives the highest expected net benefit. The ranking and relative differences between these strategies are very stable after only 100 probabilistic iterations. Unnecessary variation between strategies was removed by the following procedures:

- For each probabilistic iteration, the same patient sample was used across all of the scenarios compared. Thus the set of patients used in probabilistic loop n for the base case scenario, was the same as for probabilistic loop n for the no antithrombotic and aspirin scenarios. For probabilistic loop $n+1$, the patient sample was different from that in loop n , but again the same across all three scenarios.
- Similarly, the n 'th probabilistic iteration used the same set of values for all of the population parameters that did not differ between the scenarios. So, for example, the cost of treating a thromboembolism was the same for the n 'th probabilistic loop under the base case scenario and for the no thromboprophylaxis and aspirin scenarios. The

only parameters that differed between the scenarios for a given iteration were the cost of the antithrombotic treatment and the relative risk reductions on rates of thromboembolism and bleeds.

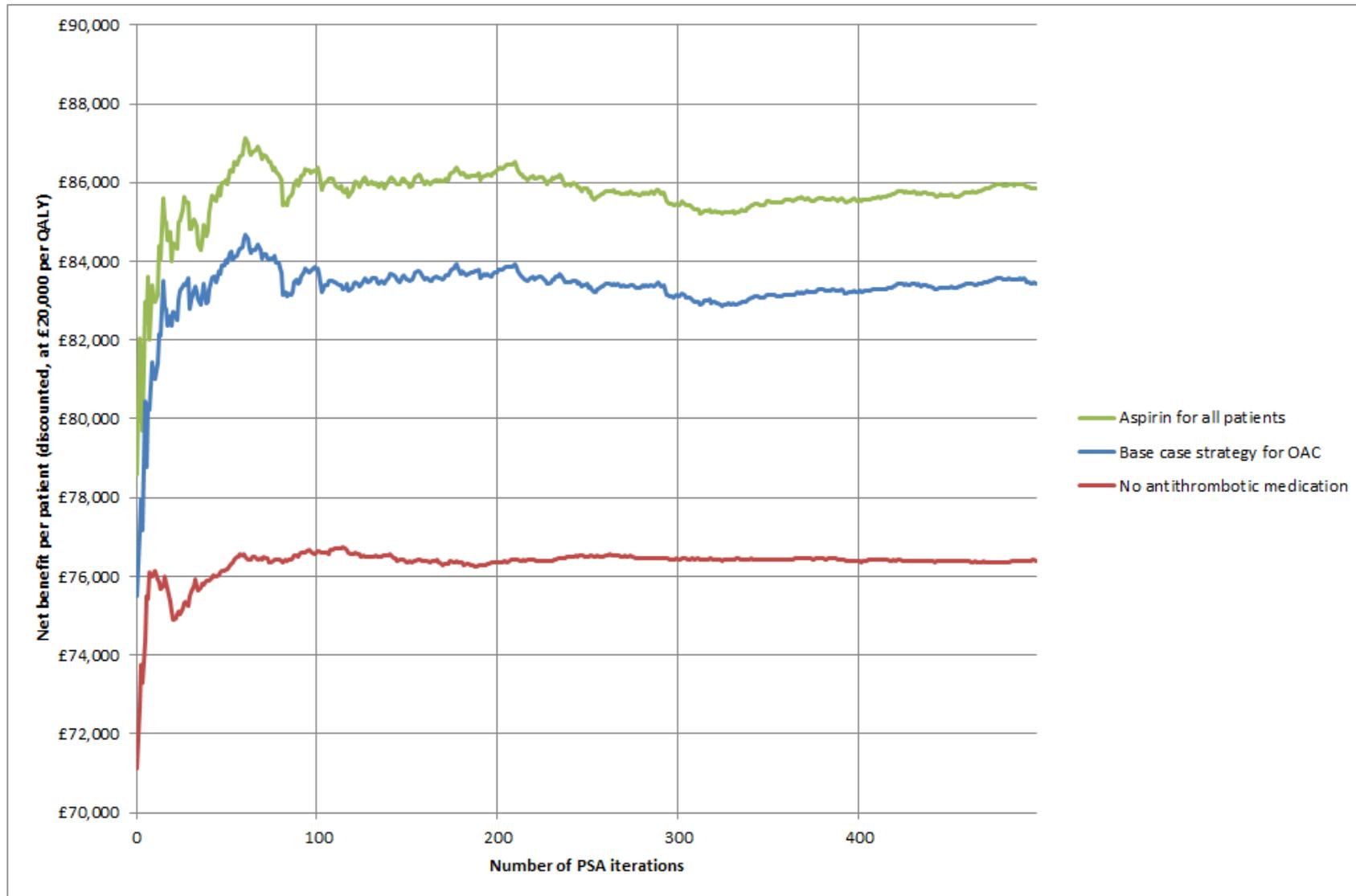
The analyses reported below used 500 probabilistic iterations with 1,000 patients per iteration.

Table 40. Base case model results (per 1,000 incident cases)

	Deterministic 1000 patients				Probabilistic 1000 patients / 500 probabilistic			
	Mean	SD	95% CI		Mean	95% CI		
Thromboembolic Events	0.31	0.71	0.24	to 0.37	0.35	0.33	to	0.36
Haemorrhagic Events	0.84	1.09	0.74	to 0.94	1.03	0.99	to	1.07
Life Years	10.68	8.97	10.12	to 11.25	10.76	10.66	to	10.85
QALYs (Undiscounted)	6.76	5.58	6.41	to 7.11	6.83	6.77	to	6.90
QALYs (Discounted at 3.5% pa)	5.20	3.51	4.98	to 5.42	5.22	5.19	to	5.26
Medication costs (Undiscounted)	5,778	6,152	5,389	to 6,167	5,787	5,684	to	5,891
Medication costs (Discounted)	4,466	4,227	4,198	to 4,733	4,386	4,317	to	4,455
Other costs (Undiscounted)	23,269	51,681	20,001	to 26,538	24,867	24,353	to	25,380
Other costs (Discounted)	15,028	28,702	13,213	to 16,843	16,662	16,315	to	17,010
Total costs (Undiscounted)	29,048	53,822	25,644	to 32,452	30,654	30,169	to	31,139
Total costs (Discounted at 3.5% pa)	19,494	29,857	17,605	to 21,382	21,048	20,729	to	21,367
Net benefit (undiscounted)*	106,176	99,704	99,871	to 112,482	106,026	104,633	to	107,419
Net benefit (discounted)*	84,497	65,099	80,380	to 88,615	83,441	82,517	to	84,365

* At £20,000 per QALY

Figure 28. Cumulative net benefit by probabilistic iteration (1,000 patients per iteration)



Topic B: PCV for patients with structural heart disease

The results of the analysis comparing amiodarone with class 1c drugs for pharmacological cardioversion in patients without structural heart disease is shown in Table 41. These suggest that the current guideline recommendation (class 1c drugs) is likely to be more cost-effective than amiodarone. Amiodarone is dominated as it gives a higher expected cost (£10 per patient) and fewer expected QALYs (-0.0018 per patient). At a cost-effectiveness threshold of £20,000 per QALY, this gave an expected incremental net benefit of £46 more per patient with class 1c drugs than with amiodarone. However, there is a high degree of uncertainty over this result. The estimated probability that class 1c drugs are more cost-effective than amiodarone is 57% at a cost-effectiveness threshold of £20,000 per QALY. The results were very similar at a cost-effectiveness threshold of £30,000 per QALY. Sensitivity analysis over the relative incidence of complications also made very little difference to these results.

Table 41. Model results: PCV drugs for patients without structural heart disease (per 1,000 patients)

	<i>Life years</i>	<i>QALYs</i>	<i>Cost (£m)</i>	<i>Net benefit (£m)</i>	<i>ICER</i>	<i>Probability cost-effective</i>
B0 Class 1c	10,758	5,225	21.044	83.453	Dominant	57%
B1 Amiodarone	10,756	5,223	21.054	83.407	Dominated	43%
INB (B1-B0)	-2.5	-1.8	0.010	-0.046		

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

INB: Incremental net benefit, compared with base case.

Topic C: Rate versus rhythm control for patients with persistent AF

The results for the illustrative comparison of rate and rhythm control strategies for patients with persistent AF are shown in Table 42.

These suggest that referring all patients with persistent AF straight to rate control dominates both the current guideline recommendations and the strategy of rhythm control. Overall, the model estimates that rate control would save £256 and yield an additional 0.017 life years and 0.133 QALYs on average per patient. The estimated differences in net benefit between the strategies are large. For example, rate control gives an expected incremental net benefit of £2,908 per patient compared with the base case at a cost-effectiveness threshold of £20,000 per QALY. This estimate increases to £4,223 at £30,000

per QALY. The model also suggests that there is a high degree of certainty over this result: in all 500 probabilistic iterations, rate control was more cost-effective than either of the other strategies. However as noted above, it is not clear that this analysis properly accounts for the full range of adverse effects associated with both rate and rhythm control drugs. Further analysis is required before conclusions should be drawn regarding the relative effectiveness or cost-effectiveness of rate and rhythm control.

Table 42. Model results: rate vs. rhythm control for persistent AF (per 1,000 incident cases)

Strategy	Life years	QALYs	Cost (£m)	Net benefit (£m)	ICER	Probability cost-effective
C0 Base case	10,758	5,224	21.05	83.44	Dominated	0%
C1 Rhythm	10,684	5,004	22.42	77.67	Dominated	0%
C2 Rate	10,775	5,357	20.79	86.35	Dominant	100%
INB (C1-C0)	-74	-220	1.37	-5.78		
INB (C2-C0)	17	133	-0.26	2.91		

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

INB: Incremental net benefit, compared with base case.

Topic E: Thromboembolism risk thresholds for oral anticoagulation

The results of the comparison of CHA₂DS₂-VASc thresholds for warfarin are shown in Table 43. For this analysis, we assumed that patients would receive Warfarin if their CHA₂DS₂-VASc score was greater than or equal to a defined threshold X and aspirin if their CHADVASC score was less than X. With X=0, all patients would be prescribed warfarin. As X is increased, fewer patients are prescribed warfarin, until with X=10 all patients are prescribed aspirin. The model assumes that patients who have a bleed while on warfarin will switch to aspirin. All other model parameters are held at their base case values. The results presented are means across 500 probabilistic iterations, each including 1,000 patients.

It can be seen that as the threshold for warfarin is increased (restricting use of anticoagulation), the mean number of thromboembolic events rises, and the mean number of bleeds falls. Both life years and QALYs are at a maximum when all patients are prescribed aspirin. This is because of the relatively high health loss associated with bleeds. However, costs are at a minimum with a CHA₂DS₂-VASc threshold of 5, and at a cost-effectiveness threshold of £20,000 per QALY gained, net benefits reach a maximum at a CHA₂DS₂-VASc threshold of 6.

The results of a similar analysis using CHADS₂ thresholds for warfarin are shown in Table 44. Here the maximum net benefit is reached at a CHADS₂ threshold of 4.

Table 43. Model results: CHA₂DS₂-VAsC thresholds for warfarin (per 1,000 incident cases)

Strategy	Threshold	Thrombo-embolisms	Bleeds	Life years	QALYs	Cost (£m)	INB (£m)
E0	Warfarin	354	1,058	10,663	5,189	20.329	0.001
E1	CHA ₂ DS ₂ -VAsC ≥ 1	355	1,059	10,677	5,193	20.318	0.109
E2	CHA ₂ DS ₂ -VAsC ≥ 2	357	1,049	10,707	5,205	20.076	0.591
E3	CHA ₂ DS ₂ -VAsC ≥ 3	365	1,033	10,766	5,229	19.932	1.201
E4	CHA ₂ DS ₂ -VAsC ≥ 4	384	994	10,827	5,252	19.668	1.936
E5	CHA ₂ DS ₂ -VAsC ≥ 5	411	952	10,885	5,272	19.640	2.369
E6	CHA₂DS₂-VAsC ≥ 6	432	925	10,924	5,284	19.732	2.510
E7	CHA ₂ DS ₂ -VAsC ≥ 7	455	896	10,943	5,289	19.875	2.459
E13	Aspirin	470	884	10,962	5,293	20.031	2.395

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

INB: Incremental net benefit, compared with base case. CV: CHA₂DS₂-VAsC

Table 44. Model results: CHADS₂ thresholds for warfarin (per 1,000 incident cases)

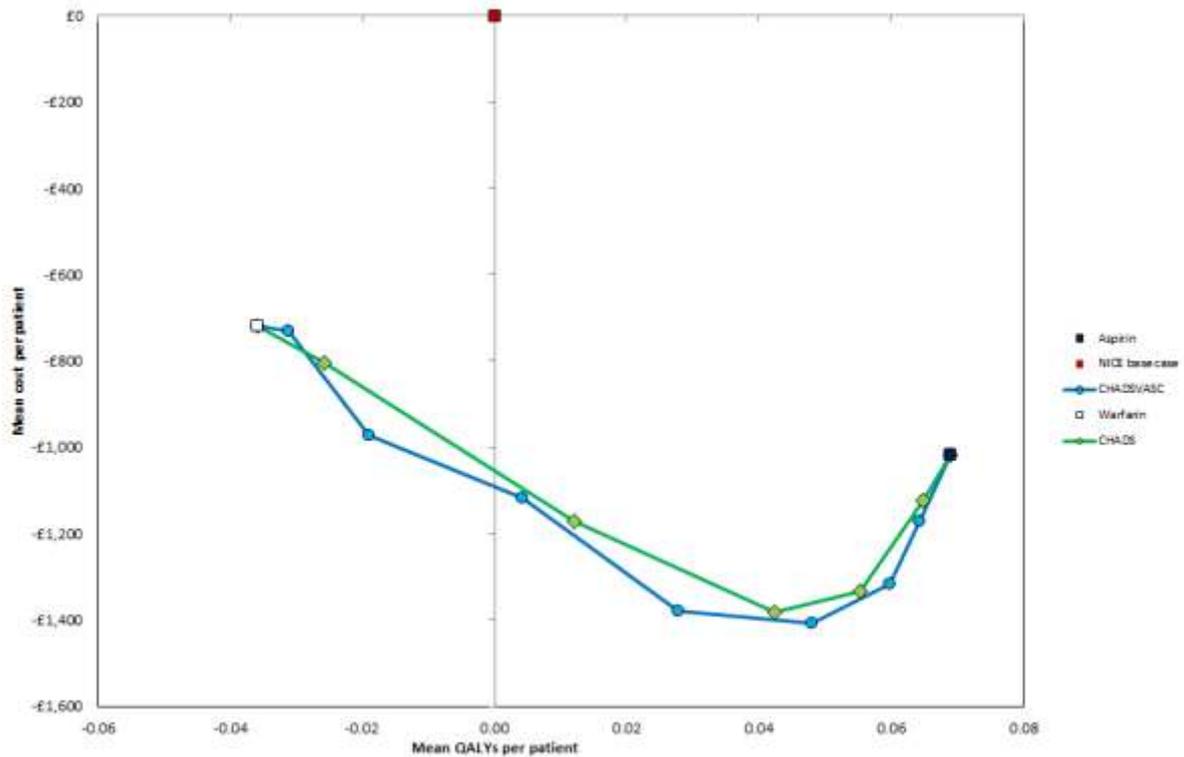
Strategy	Threshold	Thrombo-embolisms	Bleeds	Life years	QALYs	Cost (£m)	INB (£m)
E0	Warfarin	354	1,058	10,663	5,189	20.329	0.001
E8	CHADS ₂ ≥ 1	356	1,053	10,692	5,199	20.244	0.287
E9	CHADS ₂ ≥ 2	372	1,023	10,786	5,236	19.877	1.410
E10	CHADS ₂ ≥ 3	404	965	10,868	5,267	19.666	2.231
E11	CHADS₂ ≥ 4	426	936	10,911	5,280	19.715	2.440
E12	CHADS ₂ ≥ 5	453	903	10,945	5,289	19.926	2.418
E13	Aspirin	470	884	10,962	5,293	20.031	2.395

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

INB: Incremental net benefit, compared with base case.

The above results are illustrated in Figure 29. This shows estimated costs and QALYs relative to the base case results. As the CHADS₂ and CHA₂DS₂-VAsC risk thresholds are increased, the estimated QALY gain increases, and the estimated cost-effectiveness point moves to the right. This graph shows the similarity of results based on CHA₂DS₂-VAsC and on the simpler CHADS₂ scoring system.

Figure 29. Costs and effects for CHADS₂ and CHA₂DS₂-VASc thresholds for warfarin



Topic F: Bleeding risk thresholds for oral anticoagulation

Results for the HAS-BLED threshold analysis are shown in Table 45. This analysis assumes that a patient would receive warfarin only if their HAS-BLED score was less than a defined threshold γ and aspirin otherwise. As in the preceding analyses, we assumed that patients who have a bleed while on warfarin will switch to aspirin, and all other model parameters were held at their base case values. The figures presented as mean values across 500 probabilistic iterations, each including 1,000 patients.

As the HAS-BLED threshold is reduced, fewer patients receive warfarin, and the mean rate of thromboembolisms increases, while the mean rate of bleeding declines. Health outcomes (life years and QALYs) are at a maximum and costs are at a minimum when the bleeding risk threshold is set so that no patients receive warfarin (all receive aspirin).

Table 45. Model results: HAS-BLED thresholds for warfarin (per 1,000 incident cases)

Strategy	Threshold	Thrombo-embolisms	Bleeds	Life years	QALYs	Cost (£m)	INB (£m)
F0	Warfarin	354	1,058	10,663	5,189	20.329	0.001
F1	HAS-BLED < 4	376	1,025	10,710	5,206	20.420	0.258
F2	HAS-BLED < 3	419	958	10,792	5,236	20.408	0.871
F3	HAS-BLED < 2	459	899	10,897	5,271	20.255	1.718
F4	HAS-BLED < 1	470	883	10,953	5,290	20.100	2.262
F5	Aspirin	470	884	10,962	5,293	20.031	2.395

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

Topic G: Choice of oral anticoagulation drugs

For this analysis, we changed the oral anticoagulation drugs available within the modelled treatment pathway. Strategy G0 is the base case analysis, which includes CG36 criteria for allocation of patients to aspirin or warfarin, as well as TA249 criteria for access to dabigatran and TA256 for access to rivaroxaban. G1 excludes dabigatran and rivaroxaban from the treatment options.

Table 46. Model results: comparison of oral anticoagulant drugs (per 1,000 incident cases)

Strategy	Thrombo-embolisms	Bleeds	Life years	QALYs	Cost (£m)	ICER	INB (£m)
G0 Warfarin/dabigatran/ rivaroxaban/ aspirin	346	1,028	10,758	5,224	21.048	-	-
G1 Warfarin/aspirin	354	1,058	10,663	5,189	20.329	£20,038	0.001

Mean results from 500 probabilistic iterations, 1,000 patients per iteration. Cost-effectiveness threshold = £20,000 per QALY

INB: Incremental net benefit, compared with base case.

Estimated health outcomes are rather better with the addition of the new drugs: with fewer thromboembolic events and bleeds and more life years and QALYs for G0 compared with G1. However, as might be expected, costs are increased with dabigatran and rivaroxaban. At a cost-effectiveness threshold of £20,000 per QALY, this results in a very similar expected net benefit for these two strategies and the estimated probability that G0 is more cost-effective than G1 is only 45%. However, with NICE's upper threshold for cost-effectiveness (£30,000 per QALY), the expected net benefit is higher with dabigatran and rivaroxaban: £135,685 for G0 and £128,712 for G1 (51% probability that G0 is more cost-effective than G1).

Interactions: Risk thresholds and drugs for oral anticoagulation

The model results for simultaneous changes in thresholds for stroke (CHA₂DS₂-VASc) and bleeding (HAS-BLED) for allocation of patients to warfarin or aspirin are shown Table 47.

The optimum health outcome (QALY maximum) occurs with a CHA₂DS₂-VASc threshold of ≥ 2 combined with a HAS-BLED threshold of < 1 for warfarin. However, the greatest cost savings are attained with CHA₂DS₂-VASc threshold of ≥ 5 , and no HAS-BLED threshold. Overall, the model predicts a maximum incremental net benefit compared with the base case of £2,510 per patient (at £20,000 per QALY), which is achieved with a CHA₂DS₂-VASc threshold of ≥ 6 , with no HAS-BLED threshold. This finding contrasts with the separate analyses of the HAS-BLED threshold shown above in Table 45, which suggests that a very high HAS-BLED threshold would be optimum in the absence of a threshold for thromboembolic risk.

The results for a similar analysis using CHADS₂ as the stratification system for stroke and thromboembolic risk are shown in Table 48. It can be seen that the pattern of results is very similar with CHADS₂ and CHA₂DS₂-VASc, and that the optimum incremental net benefit is actually greater with CHADS₂ than with CHA₂DS₂-VASc (£2,538). This suggests that the simple CHADS₂ risk scoring system is at least as good, if not better, than the more complex alternative, CHA₂DS₂-VASc.

Table 49 and Table 50 show estimated QALYs, costs and net benefits by CHADS₂ and HAS-BLED thresholds for selection of dabigatran and rivaroxaban, respectively. These matrices illustrate how the optimal treatment thresholds can change with the anticoagulation drug: CHADS₂ ≥ 4 and HAS-BLED < 4 for warfarin; CHADS₂ ≥ 4 with no HAS-BLED threshold for dabigatran; and CHADS₂ ≥ 3 and HAS-BLED < 2 for rivaroxaban. As might be expected, there is a high level of interaction between the choice of anti-thrombotic drug, and the thromboembolic and bleeding risk thresholds for treatment.

Table 47. Model results: HAS-BLED and CHA₂DS₂-VASc thresholds for warfarin

Incremental QALYs (mean per person, discounted)

Warfarin / aspirin		CHA ₂ DS ₂ -VASc threshold							
		None	≥1	≥2	≥3	≥4	≥5	≥6	≥7
HAS-BLED threshold	None	-0.0359	-0.0311	-0.0191	0.0042	0.0278	0.0480	0.0597	0.0643
	< 4	-0.0185	-0.0147	-0.0020	0.0170	0.0420	0.0591	0.0640	0.0674
	< 3	0.0116	0.0177	0.0339	0.0470	0.0600	0.0675	0.0694	0.0686
	< 2	0.0463	0.0488	0.0598	0.0702	0.0698	0.0674	0.0687	0.0689
	< 1	0.0657	0.0690	0.0703	0.0695	0.0688	0.0689	0.0689	0.0689

Incremental cost (mean £ per person, discounted)

Warfarin / aspirin		CHA ₂ DS ₂ -VASc threshold							
		None	≥1	≥2	≥3	≥4	≥5	≥6	≥7
HAS-BLED threshold	None	-719	-731	-972	-1116	-1381	-1408	-1316	-1173
	< 4	-628	-714	-909	-1075	-1279	-1319	-1209	-1083
	< 3	-640	-729	-910	-993	-1097	-1097	-1056	-1048
	< 2	-793	-864	-1015	-993	-1031	-1057	-1028	-1017
	< 1	-948	-1030	-1038	-1014	-1018	-1017	-1017	-1017

Incremental Net Benefit (mean £ per person, discounted)

Warfarin / aspirin		CHA ₂ DS ₂ -VASc threshold							
		None	≥1	≥2	≥3	≥4	≥5	≥6	≥7
HAS-BLED threshold	None	1	109	591	1201	1936	2369	2510	2459
	< 4	258	420	868	1415	2119	2502	2489	2431
	< 3	871	1083	1588	1932	2297	2447	2444	2420
	< 2	1718	1840	2211	2397	2427	2406	2402	2395
	< 1	2262	2410	2443	2405	2393	2395	2395	2395

Mean results compared with the base case strategy.

Incremental Net benefits calculated at a cost-effectiveness threshold of £20,000 per QALY.

Table 48. Model results: HAS-BLED and CHADS₂ thresholds for warfarin

Incremental QALYs (mean per person, discounted)

Warfarin / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	-0.0359	-0.0258	0.0120	0.0424	0.0553	0.0648
	< 4	-0.0185	-0.0096	0.0270	0.0566	0.0648	0.0684
	< 3	0.0116	0.0214	0.0494	0.0653	0.0664	0.0700
	< 2	0.0463	0.0547	0.0655	0.0682	0.0690	0.0689
	< 1	0.0657	0.0703	0.0697	0.0689	0.0689	0.0689

Incremental cost (mean £ per person, discounted)

Warfarin / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	-719	-804	-1171	-1383	-1334	-1122
	< 4	-628	-766	-1104	-1287	-1241	-1082
	< 3	-640	-791	-1100	-1134	-1090	-1025
	< 2	-793	-905	-1052	-1063	-1011	-1017
	< 1	-948	-1036	-1001	-1017	-1017	-1017

Incremental Net Benefit (mean £ per person, discounted)

Warfarin / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	1	287	1410	2231	2440	2418
	< 4	258	575	1644	2420	2538	2450
	< 3	871	1219	2087	2441	2417	2425
	< 2	1718	1999	2361	2426	2391	2395
	< 1	2262	2442	2395	2395	2395	2395

Mean results compared with the base case strategy.

Incremental Net benefits calculated at a cost-effectiveness threshold of £20,000 per QALY

Table 49. Model results: HAS-BLED and CHADS₂ thresholds for dabigatran

Incremental QALYs (mean per person, discounted)

<i>Dabigatran / aspirin</i>		<i>CHADS₂ threshold</i>					
		None	≥1	≥2	≥3	≥4	≥5
<i>HAS-BLED threshold</i>	<i>None</i>	0.0594	0.0637	0.0782	0.0757	0.0741	0.0713
	< 4		0.0709	0.0795	0.0793	0.0770	0.0712
	< 3	0.0659	0.0674	0.0754	0.0737	0.0710	0.0694
	< 2	0.0652	0.0660	0.0700	0.0686	0.0690	0.0689
	< 1	0.0688	0.0691	0.0698	0.0689	0.0689	0.0689

Incremental cost (mean £ per person, discounted)

<i>Dabigatran / aspirin</i>		<i>CHADS₂ threshold</i>					
		None	≥1	≥2	≥3	≥4	≥5
<i>HAS-BLED threshold</i>	<i>None</i>	1237	827	-33	-881	-1030	-1077
	< 4		760	-114	-859	-963	-1060
	< 3	674	147	-512	-973	-1024	-1039
	< 2	-213	-593	-1004	-1030	-1012	-1017
	< 1	-785	-1034	-998	-1017	-1017	-1017

Incremental Net Benefit (mean £ per person, discounted)

<i>Dabigatran / aspirin</i>		<i>CHADS₂ threshold</i>					
		None	≥1	≥2	≥3	≥4	≥5
<i>HAS-BLED threshold</i>	<i>None</i>	-49	448	1597	2396	2513	2502
	< 4		657	1704	2444	2502	2484
	< 3	645	1202	2021	2446	2444	2427
	< 2	1517	1914	2405	2402	2392	2395
	< 1	2162	2417	2394	2395	2395	2395

Mean results compared with the base case strategy.

Incremental Net benefits calculated at a cost-effectiveness threshold of £20,000 per QALY

Table 50. Model results: HAS-BLED and CHADS₂ thresholds for rivaroxaban

Incremental QALYs (mean per person, discounted)

Rivaroxaban / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	0.0129	0.0070	-0.0008	0.0264	0.0478	0.0614
	< 4	0.0193	0.0121	0.0085	0.0349	0.0533	0.0631
	< 3	0.0460	0.0411	0.0406	0.0573	0.0680	0.0686
	< 2	0.0649	0.0611	0.0632	0.0680	0.0690	0.0689
	< 1	0.0704	0.0694	0.0697	0.0689	0.0689	0.0689

Incremental Cost (mean £ per person, discounted)

Rivaroxaban / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	5576	5381	4121	1908	520	-455
	< 4	4239	4125	2932	875	-127	-757
	< 3	1591	1496	675	-488	-967	-1018
	< 2	-468	-576	-905	-1055	-1011	-1017
	< 1	-992	-979	-1001	-1017	-1017	-1017

Incremental Net Benefit (mean £ per person, discounted)

Rivaroxaban / aspirin		CHADS ₂ threshold					
		None	≥1	≥2	≥3	≥4	≥5
HAS-BLED threshold	None	-5319	-5240	-4138	-1379	436	1683
	< 4	-3853	-3883	-2762	-176	1193	2019
	< 3	-670	-673	137	1634	2326	2389
	< 2	1765	1799	2168	2415	2391	2395
	< 1	2399	2366	2394	2395	2395	2395

Mean results compared with the base case strategy.

Incremental Net benefits calculated at a cost-effectiveness threshold of £20,000 per QALY

Discussion

Scope of the model

This chapter has presented the methods and results of a DES model developed to estimate the costs and health effects associated with the main processes of care for people with AF: diagnosis, cardioversion, antithrombotic therapy, rate and rhythm control, and ongoing monitoring. The model covered most of the service pathway in the NICE Clinical Guideline on AF, although we did not attempt to include the prevention and treatment of post-operative AF, which is a rather separate clinical question. The model does not currently include tests or interventional procedures for patients with structural heart defects, or for people with AF refractory to medical treatment (e.g. ablation or implantable devices), though these were also excluded from the scope of the original NICE guideline. As an individual-level simulation, the MAPGuide AF model reflects heterogeneity in the patient population. It contains rich information about correlated risk factors and retains information about individuals' history as they pass through the modelled pathway, which allows greater flexibility and realism in representing variations between patients.

Data sources

Overall, the model benefits from the availability of strong data sources to inform many of its key parameters. The THIN database provided individual-level data on demographic and clinical risk factors for over 12,000 individuals close to the time of their AF diagnosis. This data was derived from routine primary care data and is broadly representative of the UK population. However, there may be flaws in the recording of information in the database, as this was collected for primary patient management purposes rather than for research. We also had to make a number of assumptions to associate the recorded data with the risk factors that we required for the model.

Estimates of the rates of thromboembolism and bleeding were taken from a large population cohort study.¹⁴⁹ The applicability of this Swedish data to the UK AF population is open to question. However, as the data used in the model were stratified by risk score (CHA₂DS₂-VASC and HAS-BLED) and adjusted for estimated treatment effects of antithrombotic therapy, this would have corrected to some extent for national differences in the distribution of risk factors and differences in treatment. Other key inputs to the

model were estimates of the effects of antithrombotic and antiarrhythmic therapy, which were drawn from network meta-analyses conducted to inform recent NICE Technology Appraisals.^{136;170} These analyses provide coherent estimates of the relative impacts of the available medications in these two key areas of AF treatment, although it should be noted that they do not necessarily reflect all current evidence as we have not updated the reviews on which they were based. The cost estimates for the various interventions and outcomes along the pathway were based on standard UK sources¹⁰⁴⁻¹⁰⁶, and estimates of the costs of stroke in people with AF were also supported by an unusually large and well-conducted UK population-based study¹⁷⁶. The utility data underlying the QALY estimates may also be seen as a strength, as they came from large sample surveys using the EQ-5D instrument: starting with UK population utilities¹⁷⁸, with adjustment for AF status from an international cohort study of patients with AF¹⁷⁹, and adjustment for adverse events from a US panel of patients with a range of chronic conditions¹⁸⁰. Although their applicability to a UK population can be questioned¹⁹⁴, these sources provided consistent estimates of the utility impacts of AF control, and of the AF-related adverse effects included in the model.

There are, however, some weaknesses in the quality of data in some other areas – most notably around the accuracy of diagnostic tests for AF, the effectiveness of different methods of cardioversion and of rate control medications. Another potentially important weakness is the sparse data on the case fatality rates for haemorrhagic strokes and other major bleeds. The high estimated rate of mortality from bleeds compared with mortality from thromboembolisms (21% versus 11%) in the model, impacts on the relative cost-effectiveness of anticoagulant drugs. The analysis would be strengthened by a more robust source of data on the case-fatality rates associated with the modelled events.

Modelling anti-thrombotic treatment strategies

The model provided a good foundation for comparison of strategies for the prevention of stroke and thromboembolism. It provided stable and consistent estimates of costs and health effects for many strategies, including changes in stroke risk and bleeding thresholds for anticoagulation, and the choice of anticoagulant medication. The structure of the model offers great flexibility to model a wide variety of strategies.

As mentioned above, although the data underlying the model are of a generally good quality, there are some model inputs in particular that require further consideration before

the results should be used to inform treatment decisions. Essentially, the model weighs up the relative frequency and severity of thromboembolic and haemorrhagic events. One key factor in this equation is the mortality associated with these different events, and the case-fatality rate for haemorrhagic strokes was only based on a very small number of cases, so might not be reliable.

The model relies on two scoring systems to stratify individuals on the basis of the risk of thromboembolism (CHA₂DS₂-VASc) and bleeding (HAS-BLED). At the individual patient level, the predictive validity of these systems is limited¹⁵⁹. However, alternative scoring systems or simple pragmatic rules (such as relying on age alone) are no more likely to be correct. And across the population of people with AF, both CHA₂DS₂-VASc and HAS-BLED are strongly predictive of the related outcomes. They therefore represent the best available foundation for modelling treatment outcomes and cost-effectiveness of oral anticoagulation for people with AF.

Modelling rate versus rhythm

Adaptation of the base case model to address questions related to the choice of rate or rhythm control strategy, and the choice of particular rhythm control drugs was less successful. The key problem was that the direct trial evidence on rate versus rhythm control was incompatible with the model structure. This might be seen as a problem with the trials in this field, which were confounded with between-arm differences in rates of anticoagulant therapy and presented results using composite outcomes (such as all-cause mortality and hospitalisation) which were difficult to interpret. However, some choices made in designing the base case model did limit our ability to capture all of the potentially important impacts of rate and rhythm control strategies, and some re-programming would be required to produce a convincing evaluation of this topic.

Firstly, the model does not currently allow for any independent effect of sinus rhythm on risk of thromboembolism. It is controversial as to whether any such effect exists, although it might appear logical that it should – if a cardiac arrhythmia increases the chance of the development of an embolism, reversion to sinus rhythm might be expected to reduce this chance. However, evidence for this hypothesis is sparse. The only evidence that we found came from a multivariate analysis of data from the AFFIRM trial¹⁸⁵. This was a cox-proportional hazards regression of ‘on treatment’ data (i.e. not intention-to-treat). It

corrected for a number of covariates, including patient characteristics (age, coronary artery disease, congestive heart failure, left ventricular dysfunction, mitral valve disease, diabetes, prior stroke or TIA, and smoking) and concomitant treatments (warfarin, digoxin and AADs). The estimated hazard ratio for mortality with sinus rhythm was 0.53 (95% CI: 0.39 to 0.72). This was of a similar magnitude to the hazard ratio for warfarin use (0.50, 95% CI: 0.37 to 0.69), and use of AADs were associated with increased mortality (1.49, 95% CI: 1.11 to 2.01). The authors concluded that one hypothesis that would explain these effects is that AADs have a beneficial effect on survival through maintenance of sinus rhythm, but that this benefit might be offset by their adverse effects. However, it is difficult to draw firm conclusions from this essentially observational study, as important confounders might have been omitted. It is also difficult for us to apply this data to the model, since the results are reported in terms of an impact on all-cause mortality. However, the model could easily be modified to allow a 'what-if' sensitivity analysis on this point, by adding a relative risk multiplier for the risk of thromboembolism.

A second limitation of our analysis on the rate versus rhythm question, relates to the way in which we modelled adverse treatment effects. The model includes treatment discontinuation rates for all drugs, but it does not differentiate discontinuation due to adverse effects, or attach any health consequences or treatment costs for adverse effects other than bleeds for antithrombotic therapy and acute-onset arrhythmias. This is appropriate for the majority of adverse effects that patients tolerate and for the more minor adverse effects that might prompt treatment withdrawal, as these will not have significant or lasting consequences. However, the omission of more serious adverse effects is a problem. This could be rectified with some relatively straightforward re-programming, and data is available to inform this extension of the model. As always with adverse effect data, the large number of different types of effects that patients experience and differences in reporting present a challenge. However, data is available to estimate rates for three broad categories of adverse effects for the rate and rhythm control drugs: pro-arrhythmic events (new cardiac arrhythmias potentially provoked by anti-arrhythmic therapy), other serious adverse events and minor adverse events^{170;183}. Estimates of the QALY impacts and costs associated with these types of adverse events are also available from the economic

analysis produced by Sanofi Aventis for the NICE Technology Appraisal of Dronedarone¹⁴⁰⁻¹⁴².

A third limitation of the use of the model to compare rate and rhythm control strategies is the weakness of evidence on the effectiveness of rate control drugs. The model currently uses data from five small randomised cross-over trials reported in CG36 (55 patients in total), meta-analysed using a simple inverse-variance method. An alternative source of data was identified during the search for evidence to inform the rate versus rhythm analysis. This comes from a post hoc analysis of AFFIRM trial data¹⁸³, which estimated the effects of beta-blockers, calcium-channel blockers and digoxin alone or in combination on the achievement of adequate ventricular rate control (defined as average heart rate \leq 80 bpm, with additional criteria for maximum heart rate during exercise and on 24 hour ECG monitoring). Though observational, this dataset is larger than that included in the model (361 patients), it directly compares single drug and combination therapies, and it is presented alongside adverse effect rates. The model could be adapted to make use of this dataset, and to provide a firmer footing for comparison of rate and rhythm control strategies.

Finally, the omission of interventional techniques from the end of the rhythm control pathway in the model might have biased the results.

Comparison across the potential update topics

In the time available, we succeeded in conducting analyses for five of the eight potential update topics.

A simple two-drug comparison for Topic B supported the current recommendation for the use of class 1c drugs, rather than amiodarone, for pharmacological cardioversion in patients without structural heart disease. This finding was highly uncertain, and the estimated incremental net benefit was relatively modest (less than £50 per patient at a threshold of £20,000 per QALY). This is not unexpected, firstly because there is little evidence of any difference in effectiveness between these drugs – class 1c drugs may be faster acting than amiodarone, but overall success rates are similar, and there was no evidence of any difference in rates of recurrence. This topic is also only relevant for a subset of patients.

The analysis for Topic C was speculative, as we believe it might have omitted some important factors. Nevertheless, the findings do suggest that the balance between rate and rhythm control is potentially of high economic importance. The model suggested that an overall rate control strategy could be more effective and cost-effective than the current recommendation of selecting patients for rhythm control. The size of the estimated incremental net benefit for rate control compared with the base case strategy was £2,908 per patient (at a threshold of £20,000 per QALY).

Taken together, the analyses of Topics E and F suggest that there is also good potential for improving the cost-effectiveness of the guideline pathway by better targeting of anticoagulation therapy on the basis of stroke and bleeding risk scores. The model estimated that the optimum strategy for selecting patients for warfarin (CHADS₂ score \geq 4 and HAS-BLED score $<$ 4) would save money and improve patient outcomes, yielding an incremental net benefit of about £2,500 per patient compared with the current guideline recommendations at the £20,000 per QALY threshold.

In contrast, the model estimated more modest gains in cost-effectiveness from the use of newer oral anticoagulants rather than warfarin. Adding NICE TA recommendations for the use of dabigatran and rivaroxaban to the CG36 recommendations for warfarin made no difference to the overall net benefit at NICE's lower cost-effectiveness threshold of £20,000 per QALY. And even the upper threshold of £30,000 per QALY, a small incremental gain in net benefit of £358 per patient.

The above analyses suggest that of the topics analysed, the targeting of rhythm control therapy and of oral anti-coagulation are the highest economic priorities for inclusion in an update of the guideline, as they both show the potential for significant improvement in the net benefit of the treatment pathway. The choice of drugs for oral anticoagulation and for pharmacological cardioversion of patients without structural heart disease are relatively lower priorities from an economic perspective, as they appear to add less to overall net benefits.

Chapter 6. Discussion

Motivation for the MAPGuide project

The current approach to assessing cost-effectiveness in most clinical guidelines is partial. Some guideline developers do not explicitly take account of cost-effectiveness, and those that do often have limited health economic resources available to them.¹⁰ NICE is unusual internationally in requiring its guideline development groups to consider cost-effectiveness, and in providing resources to support them in this activity.³⁷ Each guideline has a dedicated health economist who reviews the economic literature and conducts new analyses for selected questions. This usually involves the development of a small number of separate decision models to evaluate discrete aspects of diagnosis, treatment or ongoing care, with remaining aspects being handled in a more qualitative way.

An alternative approach would be to develop a single model of the entire care pathway which is capable of providing a platform for evaluating the cost-effectiveness of multiple topics across the guideline. This has been suggested as a means of strengthening the analytical foundation of NICE clinical guidelines in order to apply the Institute's decision making principles.⁸ Several generic disease models that could fulfil such a function have been reported previously^{42;43;45;47;49}, and a methodological framework on how to design, build, check and apply 'Whole Disease Models' has been published.⁵⁰ Tappenden and colleagues⁶³ have also demonstrated how this concept could be applied to clinical guidelines by showing how their Whole Disease Model of colorectal cancer could evaluate 11 of the 15 topics addressed in the NICE colorectal cancer guideline.⁴⁰

It is uncertain, however, whether such large-scale models could be developed within the constrained timelines and resources of the NICE clinical guidelines programme, or if so whether they would provide a greater quantity or quality of cost-effectiveness evidence to support guideline recommendations than existing methods. There may also be a risks associated with the devoting all analytical resources to the development of a single complicated model. The MAPGuide project was therefore designed to further explore the feasibility and usefulness of this approach.

Summary of main findings

Feasibility of full guideline modelling

Process of model development

The project comprised of two selected case studies in which we developed models for published NICE guidelines: prostate cancer and atrial fibrillation.^{55;125} The guidelines were developed in parallel by two teams of modellers who mostly worked independently but followed an agreed protocol and met regularly to discuss their experiences and possible solutions to the problems encountered. The process of model development for both teams broadly followed that described in the methodological framework developed by Tappenden *et al.*⁵⁰ The MAPGuide models do not meet the definition of a ‘Whole Disease Model’, as they do not cover the entire breadth of the pathway from preclinical disease, diagnosis, treatment and follow-up. The process of model development was also more iterative than that described by Tappenden *et al.*, as during model development there was not a clear division between the stages of: i) problem-oriented conceptual modelling; ii) design-oriented conceptual modelling; and iii) implementation modelling. However, the models did progress through these stages, and the distinction between ‘service pathways’ and ‘disease processes’ was particularly helpful in understanding how to conceptualise the models - this contrasts with conventional health economic decision models, where this distinction is often blurred in the definition of ‘health states’.

Resource requirements

It must be acknowledged that the development of both models took longer and involved a greater input of analytical resources than was initially envisaged. In the original project plan, we estimated that the whole process from the preliminary literature review to writing up the model results would take 16 months, with a total input of 12 months of analyst time per model. In the event, the development process took closer to 24 months, and more than one whole time equivalent analyst per guideline was required. This was due to several factors. Development of the conceptual models was initially slow as the teams worked out how to articulate the models of the service pathways and disease processes. The teams also had some difficulties in understanding the intent behind some of the guideline recommendations and in converting the relatively informal guideline ‘algorithms’ into the

fully-articulated flowcharts that are required for quantitative modelling. Access to clinical advice was essential in resolving these uncertainties. To some extent, the difficulty experienced may have related to the context in which we conducted this research, which was outside of routine guideline development and in the absence of a constituted GDG. The need to assemble a large number of data inputs and to convert them into the correct format to estimate the time to event model parameters represented another challenge, as did the process of verifying and validating such large models.

Overall, there was a steep learning curve for the health economists with little experience of DES, as the structure and data requirements for DES models are quite different to those of conventional economic decision models. Consequently the teams relied more than was planned on advice and assistance from experts in simulation modelling. Conversely, for simulation modellers without experience of health economic evaluation there was also a learning process to understand how to jointly model disease processes and service pathways together, and how to incorporate epidemiological and clinical trial data into a DES format. There is currently a lack of applied texts and tutorial materials to fill this learning gap between decision modelling for economic evaluation of health care technologies and discrete event simulation. One of the strengths of the MAPGuide project was that it brought together experienced guideline developers, health economists and simulation modellers, and facilitated an exchange of ideas, knowledge and skills.

This experience might suggest that it would not be possible to develop a full guideline model within existing NICE clinical guideline development timelines and resources. However, some considerations might temper this conclusion. Many of the problems encountered in this project related to understanding the general principles of how to encode, implement and parameterise simulation models to conduct cost-effectiveness analyses within clinical guideline pathways. Having solved many of these problems for two illustrative examples, certain aspects of the development of a third full guideline model might be easier: the descriptions of the two case study models in this report provide a template for how to articulate service pathways and disease processes and the relationship between them; and some key learning points on good practice in model development are provided below. We have also developed a much better understanding of the type of expertise that is needed for the successful completion of a full guideline model. We found

that we did need specialist input in simulation modelling to complement expertise in economic evaluation and decision modelling.

In some respects model definition and parameterisation might also be easier in the context of real guideline development than in this rather artificial research context. Interpreting the thinking underlying the original guideline was often difficult, and access to the clinical expertise of a constituted GDG and the information science and systematic reviewing skills of an NCC technical team would have helped greatly. On the other hand, developing the prototype models concurrently with the 'live' guideline development would have presented other challenges, particularly the need to explain and achieve acceptance of the approach while learning how to implement it ourselves.

Scope of the models

The modelling teams succeeded in constructing individual-level simulations to predict how a heterogeneous cohort of patients with incident disease would pass through the currently-recommended pathways of care, in order to estimate key clinical outcomes, QALYs and healthcare costs. Data to inform the model parameters was obtained from the evidence reported in the original guideline, supplemented with more recent evidence where necessary. The modellers managed to cover the large majority of the breadth of the guideline pathways, if not their entirety. Exceptions included the prevention and treatment of post-operative AF, which was thought to be too different to AF not related to surgery to merge into a single model.

The modelling teams also had difficulty in representing the diagnostic sections of the pathway due to inherent problems in quantifying test accuracy for conditions lacking a diagnostic 'gold standard', the lack of good quality data on natural history (particularly for prostate cancer), and the challenge of predicting what happens to patients following a false positive or false negative test result. It is unclear whether this difficulty in modelling diagnosis was a feature of the particular guidelines chosen as case studies, or whether it reflects a more general problem in establishing diagnostic accuracy. If the latter, this would represent a limitation on the usefulness of full guideline models, although it is unlikely that conventional decision analytic models would perform any better in such circumstances. In general, one might expect full guideline models to be well suited to the evaluation of

diagnostic tests, as they are designed to capture the downstream pathways of treatment and care.¹⁹⁵

Ultimately, though, any modelling exercise requires judgement about the level of detail to be represented and the extent to which assumptions will be used to fill gaps in data. This is true for full guideline models, as for more conventional economic decision models.

Usefulness of the full guideline models

Coverage of update topics

Both modelling teams were generally successful at adapting their model to evaluate the potential update topics identified in the survey of guideline stakeholders. The prostate cancer model produced cost-effectiveness estimates for six out of the nine shortlisted topics (two of which were modelled together), and estimates of cost-effectiveness were produced for five of the eight AF topics. This represents a much better coverage of clinical questions with economic evidence than is normally possible in NICE clinical guidelines, albeit with a greater input of economic resources on the selected guidelines.

There are some real constraints on the flexibility of the models – for example, when the scope is restricted by missing information about natural history, as discussed above. Another factor that may constrain the flexibility of full guideline models to address all questions of interest is that evidence underlying different parts of the pathway may be incompatible. An example of this that arose for the AF model related to the evidence base on rate versus rhythm control strategy, which was confounded by differences in rates of anticoagulation and the use of outcomes that could not be directly incorporated in the chosen model structure (all-cause mortality). It should also be acknowledged that some of the analyses presented above are essentially illustrative, due to limitations in the data and time available within this research project. However, these analyses could be readily updated given clearer definitions of the decision problems, systematic reviews of evidence and clinical advice. The models also have the capability to address most of the topics not analysed in this report – for example, when evidence of effectiveness and a market price becomes available for some of the new drugs highlighted by stakeholders.

Overall, the case studies have shown that full guideline models can be adapted to address a wide range of questions that might arise during an update. It is also likely that the models

could be applied to other decision problems and contexts – we envisage that both of these models will evolve further over time. If full guideline models could be re-used and adapted to address a series of guideline updates and health technology assessments, this might represent a very efficient use of analytical resources. One might envisage a stock of such ‘standing models’ providing a valuable resource for NICE. However, it does remain to be shown that guideline models are sufficiently flexible to adapt to new evidence and new technologies.

Comparison of model results with survey priorities

The model results are tentative as they are not based on up-to-date systematic reviews of evidence, or informed by the expertise and experience of a GDG. However, they point to some aspects of the current guideline pathways where an update could potentially yield particularly large health gain and/or cost savings. The ‘economic priorities’ for inclusion in updates of the clinical guidelines suggested by analysis with the full guideline models differed from the priorities stated by stakeholders.

The results of the analyses conducted with the AF model identified two priority areas for an update: firstly, the related topics of the stroke risk and bleeding risk thresholds for targeting of oral anticoagulation therapy; and secondly, criteria for making the decision about the use of a rate or rhythm control strategy for patients with persistent AF. Though far from definitive, both of these analyses suggested that the current guideline recommendations might not be optimal and that there is a potential for both health improvement and NHS financial savings. In contrast, the addition of newer oral anticoagulant drugs to the treatment pathway did not significantly improve overall net benefits; essentially because they are currently priced at a level that puts them close to the NICE cost-effectiveness threshold of £20,000 to £30,000 per QALY gained. In contrast to the model results, stakeholders rated the evaluation of newer oral anticoagulants compared with warfarin as the highest priority for an update. This might possibly reflect a focus of stakeholders on effectiveness, rather than cost-effectiveness.

Similarly, although the prostate cancer model predicted that the stakeholders’ first priority for an update - the choice of surgical technique for radical prostatectomy – would be likely to improve cost-effectiveness, the magnitude of the estimated gain in net benefit was relatively modest compared with some of the other topics modelled: notably brachytherapy

with HDR or LDR external-beam radiotherapy for localised or locally advanced prostate cancer; intermittent versus continuous hormone therapy for metastatic disease; and pelvic radiotherapy with adjuvant hormone therapy for localised disease.

Interactions between topics

The key motivation for the full guideline modelling approach is to start to respond to Alan Williams' challenge to map all of the relevant 'highways and byways' of the clinical pathway, rather than to focus only on cost-effectiveness at 'particularly tricky junctions'.⁸ There are two potential benefits of taking this broader view. Firstly, the cost-effectiveness of a decision option at a specific node may change if it is evaluated in the context of different surrounding decisions. Secondly, it provides the ability to compare the magnitude of health gains, cost impacts and net benefits across decision options. Though irrelevant to cost-effectiveness, this might provide useful information to prioritise and plan for implementation of guideline recommendations.

Within the constraints of this project, we have only managed quite limited investigation of possible interactions between changes to different parts of the care pathways. In the AF chapter, we examined interactions between risk thresholds for thromboembolism and bleeding and the choice of drug for antithrombotic therapy, and (not surprisingly) found that these decisions are closely related. However, we did not manage to investigate possible interactions between anti-thrombotic and anti-arrhythmic therapies. There are reasons to suppose that the cost-effectiveness of anti-thrombotic strategies might depend on the choice of anti-arrhythmic strategy and *vice versa*, especially: contraindications to co-prescribing of drugs, notably dronedarone and dabigatran; and the possibility of independent protective effects of rhythm control against thromboembolic events.¹⁸⁵

Although we have not succeeded in examining these effects within this project, the model does offer the potential to do so. Another interesting possibility for further investigation with the AF model relates to the placing of ablative procedures to control symptoms of arrhythmia within the care pathway. Similarly, for the prostate cancer model we did not manage to examine topic interactions within this project, but there is the potential to do so.

One aspect of the evaluations presented in this report that interferes with our ability to compare and combine decision options at different points in the care pathway relates to the use of a single incident cohort. This is likely to introduce bias into the comparison of net

benefits for interventions that appear earlier or later in the pathway. As it may take some time for the cohort to reach the end of the pathway, discounting would have a greater impact on the costs and health outcomes of late interventions than on those of early interventions. In reality there is not necessarily a lag in the impact of late-pathway recommendations, as they may be implemented for prevalent cases soon after publication of the guideline. This effect will not change the qualitative results of incremental comparisons between decision options at a single point in the care pathway or the magnitude of estimated ICERs, provided that the discount rates for costs and effects are the same.¹⁹⁶ However, to estimate and compare absolute impacts of guideline recommendations and to assess the magnitude of interactions between topics, a population approach is required. This would involve starting the model with prevalent cases distributed throughout the pathway and introducing incident cases into the model as it runs; or alternatively, a run-in period can be used to build up to a steady-state distribution of patients throughout the pathway before results are collected. This was approach was adopted for the CHD policy model⁴⁵, and is quite straightforward for DES models.

Strengths and limitations of the study

In summary, this study has shown that large portions of the care pathway can be successfully modelled together within a single discrete event simulation, at least for selected NICE clinical guidelines. This was demonstrated with two contrasting case studies for two very different diseases (one cancer, one cardiovascular), underpinned by very different types and levels of evidence. These full guideline models, together with the existing published Whole Disease Model for colorectal cancer^{50;63} provide templates, which should help in the future development of similar models. The parallel development of the two models provided opportunities for learning, as the modelling teams discussed experiences and possible solutions to challenges. The involvement of a wider group of methodologists involved in the NICE clinical guidelines programme provided an understanding of the opportunities and constraints for developing and using the full guideline modelling approach in practice, which are discussed below. The primary challenge demonstrated by this project was the time and analytical resource required to develop the full guideline models. There were also some difficulties associated with missing data – for

example, the lack of evidence relating to natural history and diagnostic accuracy mentioned above.

The potential usefulness of the full guideline models for evaluation of the cost-effectiveness of a range of topics within a guideline development process was tested through the modelling of potential update topics. The separation between the researchers who chose the topics and the modelling teams meant that the latter did not know what topics they would have to evaluate until after the base case models had been designed and programmed. This provided a test of the flexibility of the models to address unexpected topics, which is a common occurrence in guideline development, despite attempts to identify economic priorities at an early stage. Overall the models performed well in these tests, providing cost-effectiveness estimates for the majority of topics considered.

However, there were some topics for which we could not identify sufficient evidence to support a meaningful economic evaluation, or where limits on the model scope or structure meant that we could not evaluate the topics within the available time. We also only managed a very limited examination of possible interactions between the cost-effectiveness of changes in different parts of the care pathways. The existence and magnitude of any such interactions is important for assessing the added value of full guideline models, compared with a more conventional piecemeal approach to modelling. The comparison of economic priorities identified by the models with stated priorities from the stakeholder survey produced some interesting results, although we were not able to complete the planned second round survey of stakeholders, so we do not know how they might have responded to these results. The main limitation of this study, however, is the rather artificial research context, which meant that we did not test the feasibility of the approach alongside real guideline development. First, we note that the guidelines used as case studies were purposively selected from a list of published NICE guidelines that were due to be updated, and that they are not necessarily representative. On the basis of our experience, we believe that the full guideline modelling approach is feasible for some NICE guidelines, but we do not believe that it would work for all guidelines. Furthermore, the decision to use existing guidelines was a pragmatic convenience, as we could start with recommended care pathways and systematic reviews of the relevant clinical evidence. Although it would have been desirable to test the application of the approach alongside the development of a new

guideline, this was thought to be premature, as we could not be confident that the modelling teams would succeed in developing the model in time to inform the development of guideline recommendations. There were some disadvantages to conducting this study outside of 'live' guideline development. For example, the modelling teams had more limited access to clinical expertise than would usually be available to guideline economists, who can consult with clinical leads and other members of the GDG. This might have influenced the assumptions used to interpret and link the guideline recommendations. Access to the GDG and liaison with other members of the technical team should also help in identification of relevant evidence.

Value of information (VOI) analysis has become a fairly standard adjunct to economic evaluations of health care technologies. Statistics such as the Expected Value of Perfect Information (EVPI), the Expected Value of Partial Perfect Information (EVPPPI), and the Expected Value of Sample Information (EVSPI), can be used to inform decisions about the collection of further information and the rational prioritisation of research budgets.⁶⁵ In this report we chose not to present VOI estimates for two key reasons. First, given the relatively informal methods that we used to source evidence for the modelling exercises, we were not confident that we had fully characterised the uncertainty surrounding the estimates of cost-effectiveness. In this situation, the robustness of VOI estimates would be questionable. Second, it is unclear that such estimates would provide an indication of priority for updating, which was the focus of our analyses. The decision to include a topic in an update of a guideline is broadly dependent on two key factors: the perceived likelihood that the recommendation could change following a systematic review and GDG debate; and the relative importance of any such change (expected impact on population health and health care expenditure). A high EVPI indicates that a decision is both important and uncertain, however one would not generally decide to include such a decision in a guideline update unless there was a reasonable expectation that a systematic review and consideration by a panel of experts could help to resolve the uncertainty. There is therefore not a clear relationship between EVPI and priority for a guideline update.

This is not to say that VOI methods are not potentially of value in guideline development, as they might help GDGs to decide on research recommendations. It is also possible that an EVPPPI analysis - in which the relative contributions of different model parameters, or groups

of parameters, to EVPI is quantified - could help to inform decisions about searching for better information to improve a full guideline model. However, there are practical difficulties in estimating EVPPI for a DES model, as the conventional approach requires an additional level of iteration which would considerably increase model run time.¹⁹⁷

Implications for guideline development

Pathway versus piecemeal models

A number of claims may be made for the advantages of the full guideline modelling approach investigated in this report compared with the current approach to economic evaluation in NICE clinical guidelines. The full guideline models that we developed provide a framework for addressing a range of cost-effectiveness questions using a consistent set of methods, assumptions and evidence. They allow assessment of whether and how interventions in one part of the pathway influence other parts of the pathway: capturing upstream and downstream systemic effects. Once developed, the models are a resource that could be reused for future guideline updates or adapted for other economic evaluations. There may also be potential spinoff benefits for general guideline development, as modelling enforces greater clarity over the pathway, and may help to elucidate gaps and ambiguities in the existing evidence base.

There are, however, constraints on the routine adoption of this approach in the NICE clinical guidelines programme. The most obvious is the time and resources that are required. Although it is possible that learning from this project could enable faster development of full guideline models in the future, this would still be difficult within the current timelines and economic resources available to the National Collaborating Centres.

The two case studies presented in this report were selected from a list of possibilities largely on the basis that they were thought to be amenable to the full guideline modelling approach. It is unlikely that this approach would work for all clinical guidelines; for some topics the current understanding of the service pathway or disease process may be too poor, or data may be too sparse to allow credible modelling of the full guideline. It should be noted that in such cases, conventional 'piecemeal' modelling is also likely to be challenging. When it is feasible, the full guideline modelling does have the capacity to broaden the range and to improve the consistency of cost-effectiveness analysis within NICE

clinical guidelines. The case studies presented in this report have demonstrated this potential, and further consideration of the approach is warranted.

Individual level simulation versus cohort models

Another lesson from this project is the potential value of discrete event simulation for modelling the complex care pathways in clinical guidelines. Although simplicity is to be valued in modelling, “more complex areas require models that respect complexity”⁵¹. In individual-level simulation models patients carry information with them, so the model can keep track of varied, complex and evolving patterns of risk factors, clinical histories and co-morbidities as patients move through a complicated care pathway. This provides the flexibility to tailor decisions to a person’s characteristics, and to model the resulting outcomes in a more realistic way. Models can also be illustrated with more natural representations of care pathways, which are likely to be more accessible for GDGs and stakeholders than ‘twiggy’ decision trees or Markov models with multiple health states. Another useful feature of DES is it can be readily extended from the single cohort approach (taking a group of patients from some defined starting point and following them through to death) to model whole populations of prevalent and incident cases. This population approach would facilitate estimation of cost impact alongside cost-effectiveness within the same model, to provide NHS budget holders and policy makers with better estimates to inform implementation and planning. Whilst this could be achieved with more conventional decision analytic models¹⁹⁶, capturing this level of complexity in a Markov or decision tree framework would be cumbersome to develop and hard to understand.

The main drawback to adopting a DES approach to modelling in NICE guidelines would be the need for investment to develop (or to buy in) specialised skills. Most health economists do not have applied experience of using DES, and in this project we did find that it took time to acquire the necessary knowledge and understanding. We also had specialised experts in simulation modelling who worked alongside health economists with experience of more conventional cost-effectiveness modelling techniques, which worked well. Furthermore, although specialist software is not essential for simulation modelling, it does make it easier. Thus an investment in software would be necessary (in terms of money and learning time). Investment in hardware might also be necessary to keep model runtime manageable, as models that combine individual-level simulation with probabilistic sensitivity analysis can be

slow to run. We did not find this to be a limiting factor in our case studies, although we did have access to a number of fast PCs to run each model.

It is sometimes thought that DES requires more data than Markov models or decision trees. However, data requirements are more a function of model size and complexity than technique.^{51;53} There are some differences though in the type of data required, and modellers would need to learn how to identify and fit data to inform time-to-event estimates. This might also present a challenge for systematic reviewers and other guideline developers, although this will depend on the topic area and familiarity with survival analysis. Communication to ensure understanding and agreement of methods and results within the GDG is essential.

Access to individual-level data on patient characteristics at model entry is very useful (if not absolutely essential) for DES modelling, as it can build in rich correlations between risk factors. Access to THIN data for the AF model strengthened the ability of the model to reflect variation between patients. In the absence of a data source for key model parameters, calibration can be used to infer missing or unknowable data, as in the prostate cancer model. In the absence of routine sources of individual patient data for a topic, suitable data might sometimes be available from disease-specific clinical audit or registry databases. More generally, any modelling approach needs to make use of the best available evidence at the time of analysis.

Good practice in model development

Due to the size and complexity of full guideline models, observation of good practice in model design, implementation and verification/validation is particularly crucial. Some key issues that arose in the case studies are discussed below:

- The need for clarity about the boundaries of the model. This should generally reflect the guideline scope, but it might be extended where there are spill over effects from out of scope issues on the cost-effectiveness of guideline topics or *vice versa*. For example, sometimes guidelines include referral for specialist assessment but not the subsequent specialist treatment (e.g. AF guideline tertiary referral and ablation). There may also be circumstances where parts of the pathway included in the scope are difficult to model.

For example, in both of our case studies, we found it difficult to model early case identification due to lack of data on natural history and diagnostic accuracy.

- The need for clarity over whether the model pathway is meant to reflect recommended practice (i.e. an existing guideline pathway) or current practice. For our case studies we aimed to model the recommended pathway from the current guideline. However, we had to supplement this with assumptions about current practice (as advised by clinical experts). During scoping it may be appropriate to describe more than one pathway reflecting variations in recommended or actual practice.
- It is essential to focus not just on the service pathway, but also to develop a model of the disease process (how individuals' health indicators and status progress over time). It is also essential to understand how disease and service pathways interact with one another. Discussion with experts is essential to understand pathways and disease processes. Other useful parallel sources to inform model design are existing models, observational studies and effectiveness data (which define the important and available outcome measures).
- Visual representation is extremely important in articulating pathway and disease models, both simple schematic overviews and detailed flowcharts. However, these are not sufficient – textual descriptions of the pathway and disease models may help clinical experts, GDG members and stakeholders to understand and critique models.
- As in any model, simplifying assumptions are essential, and these can restrict possible future uses of the model. The art is in deciding at what level of resolution to reflect the different stages of the pathway. This ultimately reflects the series of choices made during model development; these judgments and their implications should be clearly articulated and justified in light of the availability of evidence.
- There may be inconsistencies between different bodies of evidence that inform different sections of the model. For example, the evidence base on antithrombotic therapy for patients with AF presented results for the outcomes of thromboembolic and haemorrhagic events, and did not separate these into fatal and non-fatal events. This made it impossible to model effects on all-cause mortality. However, the evidence comparing rate and rhythm control focussed on the combined effects of treatment

through all-cause mortality, and all-cause hospitalisation. Similarly, in the prostate cancer case study, the lack of evidence relating to the joint trajectories of Prostate Specific Antigen (PSA), Gleason score and disease progression, and their impact on treatment decisions meant that it was not possible to fully reflect the way in which clinicians use this information to make decisions about individual patients.

Recommendations for research

- 1) The case study models have been made available to the NCC teams who are now developing the updates of the atrial fibrillation and prostate cancer guidelines. This provides an opportunity to observe the impact and perceived usefulness of the models within the NICE clinical guidelines programme. Research should be conducted to seek the opinions of the members of the NCC technical team, the GDGs, stakeholders and the NICE guidelines team to determine whether they made use of the models, and if so whether the results were useful in informing guideline recommendations.
- 2) Further development of the case study models to assess the existence and magnitude of possible interactions between changes to different parts of the care pathways would be informative. The existence of sizeable interactions is a crucial element in judging the value of full guideline modelling compared with more conventional partial evaluation of sections of the care pathway.
- 3) Another useful development of the case study models would be to extend them to estimate budget and health impacts across a whole patient population (rather than for single incident cohort). Currently, NICE conducts cost impact assessments for clinical guidelines separately from the cost-effectiveness analyses conducted by NCCs, which may lead to inconsistent estimates. Furthermore, extending the models to a population perspective would enable more robust comparison of the net benefit of changes at different points in the care pathway without artificial distortion from discounting.
- 4) The next step would be to apply the full guideline modelling approach to a new NICE guideline. This would need additional resources to support the analytical effort, for example funding for a simulation modeller to work with the NCC economist on the model development. Additional challenges to be faced would include: how to develop the initial understanding of the care pathway through elicitation from GDG members, stakeholders or other experts, particularly in areas where there are important variations in current practice; how to work

with the GDG in developing, using and interpreting the guideline model; and how to communicate the methods and underlying assumptions to stakeholders.⁴) There are some areas where further development of methods would be useful:

- a. Methods for eliciting expert opinion and reaching consensus about the structure of disease process and service pathway models to inform guideline development and economic modelling;
- b. Methods for robust model calibration to infer missing or unobservable parameters in complex decision models;
- c. Development of standardised software templates or methods of presentation to help guideline economists to develop flexible and accessible full guideline models in consultation with other guideline methodologists, GDGs and stakeholders;
- d. Methods to test the internal and external validity of full guideline models.

References

- (1) National Institute for Health and Clinical Excellence. The management of male lower urinary tract symptoms (LUTS). NICE; 2010.
- (2) National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. London: NICE; 2008.
- (3) Institute of Medicine. Guidelines for clinical practice: from development to use. Washington DC: National Academic Press; 1992.
- (4) Institute of Medicine. Clinical practice guidelines we can trust. 2011.
- (5) Qaseem A, Forland F, Macbeth F, Ollenschl+ñger G+, Phillips S, van der Wees P et al. Guidelines International Network: Toward International Standards for Clinical Practice Guidelines. *Ann Intern Med* 2012; 156(7):525-531.
- (6) Eddy DM. A manual for assessing health practices and designing practice policies: the explicit approach. Philadelphia: American College of Physicians; 1992.
- (7) Williams A. How should information on cost effectiveness influence clinical practice? In: Delamothe T, editor. Outcomes into clinical practice. London: BMJ Publishing Group; 1995. 99-107.
- (8) Williams A. What could be nicer than NICE? London: Office of Health Economics; 2004.
- (9) Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Third ed. UK: Oxford University Press; 2005.
- (10) Edejer TTT. Improving the use of research evidence in guideline development: II. Incorporating considerations of cost-effectiveness, affordability and resource implications. *Health Research Policy and Systems* 2008; 4:23-26.
- (11) Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. 1st ed. Oxford: Oxford University Press; 2006.
- (12) Eccles M, Mason J. How to develop cost-conscious guidelines. *Health Technol Assess* 2001; 5(16).
- (13) Culyer AJ. NICE's use of cost effectiveness as an exemplar of a deliberative process. *Health Economics Policy and Law* 2006; 1:299-318.
- (14) Thomson R, Parkin D, Eccles M, Sudlow M, Robinson A. Decision analysis and guidelines for anticoagulant therapy to prevent stroke in patients with atrial fibrillation. *The Lancet* 2000; 355(9208):956-962.

- (15) Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J et al. Grade guidelines: 10. Considering resource use and rating the quality of economic evidence. *Journal of Clinical Epidemiology* 2012; In press.
- (16) Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A et al. GRADE: incorporating considerations of resource use into grading recommendations. *Br Med J* 2008; 336:1170-1173.
- (17) Banta D. The development of health technology assessment. *Health policy (Amsterdam, Netherlands)* 2003; 63(2):121-132.
- (18) Cochrane AL. Effectiveness and efficiency. Random reflections on health services. Cambridge: Cambridge University Press; 1972.
- (19) Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: A six-country comparison. 1420 Vol 91 ed. Commonwealth Fund; 2010.
- (20) Neumann PJ. Lessons for Health Technology Assessment: It Is Not Only about the Evidence. *Value in Health* 2009; 12:S45-S48.
- (21) Coast J. Is economic evaluation in touch with society's health values? *BMJ* 2004; 329(7476):1233-1236.
- (22) Buxton MJ, Drummond MF, van Hout BA, Prince RL, Sheldon TA, Szucs T et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997; 6(3):217-227.
- (23) Sheldon TA. Problems of using modelling in the economic evaluation of health care. *Health Econ* 1996; 5(1):1-11.
- (24) 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).
- (25) Chilcott J, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Paisley S et al. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. *Health Technol Assess* 2010; 14(25).
- (26) Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health* 2012; 15(6):843-850.
- (27) Anderson R. Systematic reviews of economic evaluations: utility or futility? *Health Econ* 2010; 19(3):350-364.
- (28) National Institute for Health and Clinical Excellence. Framework document. London: NICE; 2004.

- (29) Chidgey J, Leng G, Lacey T. Implementing NICE guidance. *JRSM* 2007; 100(10):448-452.
- (30) National Institute for Health and Clinical Excellence. How NICE clinical guidelines are developed: an overview for stakeholders, the public and the NHS. London: NICE; 2009.
- (31) National Institute for Health and Clinical Excellence. The guidelines manual. London: NICE; 2009.
- (32) National Institute for Health and Clinical Excellence. Social value judgements. Principles for the development of NICE guidance. Second edition. London: NICE; 2008.
- (33) Wailoo A, Roberts J, Brazier J, McCabe C. Efficiency, equity, and NICE clinical guidelines. *BMJ* 2004; 328:536-537.
- (34) Eccles M. NICE clinical guidelines. Health economics must engage with complexity of issues. *BMJ* 2004; 329:572.
- (35) Littlejohns P, Leng G, Culyer A, Drummond M. NICE clinical guidelines. Maybe health economists should participate in guideline development. *Br Med J* 2004; 329:571.
- (36) Pilling S. NICE clinical guidelines. Account of guideline development was inadequate. *BMJ* 2004; 329:571-572.
- (37) Wonderling D, Sawyer L, Fenu E, Lovibond K, Laramée P. National Clinical Guideline Centre Cost-Effectiveness Assessment for the National Institute for Health and Clinical Excellence. *Ann Intern Med* 2011; 154(11):758-765.
- (38) Thornton J, Alderson P, Tan T, Turner C, Latchem S, Shaw E et al. Introducing GRADE across the NICE clinical guideline program. *Journal of Clinical Epidemiology* 2012; In press(0).
- (39) Tappenden P, Chilcott J, Brennan A, Pilgrim H. Systematic review of economic evidence for the detection, diagnosis, treatment, and follow-up of colorectal cancer in the United Kingdom. *International Journal of Technology Assessment in Health Care* 2009; 25(04):470-478.
- (40) National Collaborating Centre for Cancer. Colorectal cancer: the diagnosis and management of colorectal cancer: Full Guideline. Cardiff: National Collaborating Centre for Cancer; 2011.
- (41) Schlessinger L, Eddy DM. Archimedes: a new model for simulating health care systems—the mathematical formulation. *Journal of Biomedical Informatics* 2002; 35(1):37-50.

- (42) Eddy DM, Schlessinger L. Archimedes: A trial-validated model of diabetes. *Diabetes Care* 2003; 26(11):3093-3101.
- (43) Clarke PM, Gray AM, Briggs A, Farmer AJ, Fenn P, Stevens RJ et al. A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2004; 47(10):1747-1759.
- (44) Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006; 15:1295-1310.
- (45) Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality and cost: The coronary heart disease policy model. *American Journal of Public Health* 1987; 77:1417-1426.
- (46) Weinstein MC. Methodologic issues in policy modeling for cardiovascular disease. *J Am Coll Cardiol* 1989; 14(3, Supplement A):A38-A43.
- (47) Cooper K, Davies R, Roderick P, Chase D, Raftery J. The Development of a Simulation Model of the Treatment of Coronary Heart Disease. *Health Care Manag Sci* 2002; 5(4):259-267.
- (48) Davies R, Roderick P, Raftery J. The evaluation of disease prevention and treatment using simulation models. *European Journal of Operational Research* 2003; 150(1):53-66.
- (49) Pilgrim H, Tappenden P, Chilcott J, Bending M, Trueman P, Shorthouse A et al. The costs and benefits of bowel cancer service developments using discrete event simulation. *J Oper Res Soc* 2008; 60(10):1305-1314.
- (50) Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole Disease Modeling to Inform Resource Allocation Decisions in Cancer: A Methodological Framework. *Value in Health* 2012; In press.
- (51) Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research & Policy* 2004; 9(2):110-118.
- (52) Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ* 2003; 12(10):837-848.
- (53) Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. *J Oper Res Soc* 2006; 58(2):168-176.
- (54) Longworth L, Bojke L, Tosh J, Sculpher M. MRC-NICE Scoping Project: Identifying the National Institute for Health and Clinical Excellence's methodological research priorities and an initial set of priorities. CHE Research Paper 51 ed. University of York: Centre for Health Economics; 2009.

- (55) National Institute for Health and Clinical Excellence. Prostate cancer diagnosis and treatment. NICE clinical guideline 58. London: NICE; 2008.
- (56) National Institute for Health and Clinical E. Atrial fibrillation: the management of atrial fibrillation: NICE clinical guideline 36. Quick reference guideline. United Kingdom: National Institute for Health and Clinical Excellence (NICE); 2006.
- (57) Kaltenthaler E, Tappenden P, Paisley S, Squires H. Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. 13 ed. DSU, SchARR, University of Sheffield; 2011.
- (58) Tappenden P. Conceptual modelling for health economic model development. 12.05 ed. University of Sheffield; 2012.
- (59) Law AM. Simulation modeling and analysis. Fourth ed. McGraw-Hill; 2007.
- (60) Latmer N. Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 14 ed. Sheffield: Decision Support Unit, SchARR, University of Sheffield; 2011.
- (61) Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. 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. *Health Technol Assess* 2004; 8(11).
- (62) Brailsford SC, Bolt TB, Bucci G, Chausalet TM, Connell NA, Harper PR et al. Overcoming the barriers: a qualitative study of simulation adoption in the NHS. *J Oper Res Soc* 2011.
- (63) Tappenden P, Brennan A, Chilcott J, Squires H. PCN193 Using Whole Disease Modelling to Inform Economic Recommendations for the Detection, Diagnosis, Treatment and Follow-up of Colorectal Cancer. *Value in Health* 2011; 14(7):A469-A470.
- (64) Koerkamp BG, Stijnen T, Weinstein MC, Hunink MGM. The Combined Analysis of Uncertainty and Patient Heterogeneity in Medical Decision Models. *Med Decis Mak* 2011; 31(4):650-661.
- (65) Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ* 2005; 14(4):339-347.
- (66) Doubilet PM, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. *Med Decis Mak* 1985; 5(2):157-177.
- (67) Cancer Research UK. Prostate cancer statistics. 2012. 4-12-0012.
- (68) Prostate Cancer Risk Management Programme. Information about prostate cancer. 2012. 4-12-0012.

- (69) Chilcott J, Hummel S, Mildred M. Option appraisal: screening for prostate cancer. Report to the UK National Screening Committee. University of Sheffield: School of Health and Related Research (ScHARR); 2010.
- (70) National Collaborating Centre for Cancer. Prostate cancer: diagnosis and treatment. Full Guideline. Cardiff: National Collaborating Centre for Cancer; 2008.
- (71) National Collaborating Centre for Cancer. Prostate cancer: diagnosis and treatment. Evidence review. Cardiff: National Collaborating Centre for Cancer; 2008.
- (72) National Institute for Health and Clinical Excellence. Prostate cancer diagnosis and treatment. Quick reference guide. London: NICE; 2008.
- (73) National Institute for Health and Clinical Excellence. Prostate cancer diagnosis and treatment. Implementation advice. 3rd edition ed. London: NICE; 2011.
- (74) National Institute for Health and Clinical Excellence. Prostate cancer diagnosis and treatment. Scope. London: NICE; 2005.
- (75) National Institute for Health and Clinical Excellence. Referral guidelines for suspected cancer. NICE Clinical Guideline 27. London: NICE; 2005.
- (76) National Institute for Health and Clinical Excellence. Docetaxel for the treatment of hormone refractory prostate cancer. Technology appraisal TA101. London: NICE; 2006.
- (77) Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C et al. Radical prostatectomy versus watchful waiting in early prostate cancer. 2011; 2011/05/06:1708-1717.
- (78) Prostate Cancer Risk Management Programme. Prostate Cancer Risk Management Programme information for primary care; PSA testing in asymptomatic men. Evidence document. 2012. 4-12-0012.
- (79) Cancer Intervention and Surveillance Modeling Network. Prostate Cancer Model Profiles. 2012. CISNET.
- (80) D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. 1998; 1998/09/28:969-974.
- (81) Tilling K, Garmo H, Metcalfe C, Holmberg L, Hamdy FC, Neal DE et al. Development of a new method for monitoring prostate-specific antigen changes in men with localised prostate cancer: a comparison of observational cohorts. 2010; 2009/03/24:446-452.

- (82) Bosch J, Tilling K, Bohnen AM, Donovan JL. Establishing normal reference ranges for PSA change with age in a population-based study: The Krimpen study. *Prostate* 2006; 66(4):335-343.
- (83) Sakr WA, Haas GP, Cassin BF, Pontes JE, Crissman JD. The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *The Journal of urology* 1993; 1993/08/01:379-385.
- (84) Sanchez-Chapado M, Olmedilla G, Cabeza M, Donat E, Ruiz A. Prevalence of prostate cancer and prostatic intraepithelial neoplasia in Caucasian Mediterranean males: an autopsy study. *The Prostate* 2003; 2003/01/09:238-247.
- (85) Shiraishi T, Watanabe M, Matsuura H, Kusano I, Yatani R, Stemmermann GN. The frequency of latent prostatic carcinoma in young males: the Japanese experience. *In vivo (Athens, Greece)* 1994; 1994/05/01:445-447.
- (86) Office for National Statistics. UK Interim life tables, 1980-82 to 2008-10. ONS; 2011.
- (87) Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. 1995; 49:327-335.
- (88) Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Medical decision making : an international journal of the Society for Medical Decision Making* 2011; 2010/12/04:625-641.
- (89) Rabbani F, Stroumbakis N, Kava BR, Cookson MS, Fair WR. Incidence and clinical significance of false-negative sextant prostate biopsies. 1998; 159:1247-1250.
- (90) Raaijmakers R, Kirkels WJ, Roobol MJ, Wildhagen MF, Schrder FH. Complication rates and risk factors of 5802 transrectal ultrasound-guided sextant biopsies of the prostate within a population-based screening program. *Urology* 2002; 2002/11/14:826-830.
- (91) Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L et al. Prostate Testing for Cancer and Treatment (ProtecT) feasibility study. *Health technology assessment (Winchester, England)* 2003; 2003/04/24:1-88.
- (92) Alibhai SM, Leach M, Tomlinson G, Krahn MD, Fleshner N, Naglie G. Rethinking 30-day mortality risk after radical prostatectomy. *Urology* 2006; 2006/11/23:1057-1060.
- (93) Fransson P, Lund JA, Damber JE, Klepp O, Wiklund F, Fossa S et al. Quality of life in patients with locally advanced prostate cancer given endocrine treatment with or without radiotherapy: 4-year follow-up of SPCG-7/SFUO-3, an open-label, randomised, phase III trial. 2009; 2009/03/17:370-380.

- (94) Widmark A, Klepp O, Solberg A, Damber JE, Angelsen A, Fransson P et al. Endocrine treatment, with or without radiotherapy, in locally advanced prostate cancer (SPCG-7/SFUO-3): an open randomised phase III trial. *Lancet* 2009; 2008/12/19:301-308.
- (95) Calais da Silva FE, Bono AV, Whelan P, Brausi M, Marques Queimadelos A, Martin JA et al. Intermittent androgen deprivation for locally advanced and metastatic prostate cancer: results from a randomised phase 3 study of the South European Urooncological Group. 2009; 2009/03/03:1269-1277.
- (96) Eisenberger MA, Blumenstein BA, Crawford ED, Miller G, McLeod DG, Loehrer PJ et al. Bilateral Orchiectomy with or without Flutamide for Metastatic Prostate Cancer. 1998; 339:1036-1042.
- (97) Seidenfeld J, Samson DJ, Hasselblad V, Aronson N, Albertsen PC, Bennett CL et al. Single-therapy androgen suppression in men with advanced prostate cancer: a systematic review and meta-analysis. *Ann Intern Med* 2000; 2000/04/01:566-577.
- (98) Tyrrell CJ, Kaisary AV, Iversen P, Anderson JB, Baert L, Tammela T et al. A randomised comparison of 'Casodex' (bicalutamide) 150 mg monotherapy versus castration in the treatment of metastatic and locally advanced prostate cancer. *European urology* 1998; 1998/06/27:447-456.
- (99) Suzuki H, Okihara K, Miyake H, Fujisawa M, Miyoshi S, Matsumoto T et al. Alternative nonsteroidal antiandrogen therapy for advanced prostate cancer that relapsed after initial maximum androgen blockade. *The Journal of urology* 2008; 2008/07/19:921-927.
- (100) Venkitaraman R, Thomas K, Huddart RA, Horwich A, Dearnaley DP, Parker CC. Efficacy of low-dose dexamethasone in castration-refractory prostate cancer. *BJU international* 2008; 2007/10/19:440-443.
- (101) Petrylak DP, Tangen CM, Hussain MH, Lara PN, Jr., Jones JA, Taplin ME et al. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *The New England journal of medicine* 2004; 2004/10/08:1513-1520.
- (102) Hummel S, Simpson EL, Hemingway P, Stevenson MD, Rees A. Intensity-modulated radiotherapy for the treatment of prostate cancer: a systematic review and economic evaluation. 2010; 2010/10/30:1-108, iii.
- (103) Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K et al. 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. *Health Technol Assess* 2007; 11:iii-iv, xv.
- (104) Department of Health. 2010-11 reference costs publication. Department of Health; 2011.

- (105) Curtis L. Unit Costs of Health and Social Care 2011. University of Kent, Canterbury: Personal Social Services Research Unit (PSSRU); 2011.
- (106) Joint Formulary Committee. British National Formulary (BNF) 64. The Pharmaceutical Press; 2012.
- (107) National Institute for Health and Clinical E. Review consultation document. Review of Clinical Guideline CG58. United Kingdom: National Institute for Health and Clinical Excellence (NICE); 2011.
- (108) Matzinger O, Poortmans P, Giraud JY, Maingon P, Budiharto T, van den Bergh AC et al. Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: dummy run and individual case review. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 2009; 2008/11/29:285-290.
- (109) Martis G, Diana M, Ombres M, Cardi A, Mastrangeli R, Mastrangeli B. Retropubic versus perineal radical prostatectomy in early prostate cancer: Eight-year experience. *Journal of surgical oncology* 2007; 95(6):513-518.
- (110) Guazzoni G, Cestari A, Naspro R, Riva M, Centemero A, Zanoni M et al. Intra- and peri-operative outcomes comparing radical retropubic and laparoscopic radical prostatectomy: results from a prospective, randomised, single-surgeon study. *European urology* 2006; 2006/03/28:98-104.
- (111) Asimakopoulos AD, Pereira Fraga CT, Annino F, Pasqualetti P, Calado AA, Mugnier C. Randomized comparison between laparoscopic and robot-assisted nerve-sparing radical prostatectomy. 2011; 8:1503-1512.
- (112) Ramsay C, Pickard R, Robertson C, Close A, Vale L, Armstrong N et al. Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol Assess* 2012; 16:1-313.
- (113) Sathya JR, Davis IR, Julian JA, Guo Q, Daya D, Dayes IS et al. Randomized trial comparing iridium implant plus external-beam radiation therapy with external-beam radiation therapy alone in node-negative locally advanced cancer of the prostate. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005; 2005/02/19:1192-1199.
- (114) Hoskin PJ, Motohashi K, Bownes P, Bryant L, Ostler P. High dose rate brachytherapy in combination with external beam radiotherapy in the radical treatment of prostate cancer: initial results of a randomised phase three trial. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 2007; 2007/05/29:114-120.
- (115) Stromberg JS, Martinez AA, Horwitz EM, Gustafson GS, Gonzalez JA, Spencer WF et al. Conformal high dose rate iridium-192 boost brachytherapy in locally advanced

- prostate cancer: superior prostate-specific antigen response compared with external beam treatment. *The cancer journal from Scientific American* 1997; 1997/12/24:346-352.
- (116) Kestin LL, Martinez AA, Stromberg JS, Edmundson GK, Gustafson GS, Brabbins DS et al. Matched-pair analysis of conformal high-dose-rate brachytherapy boost versus external-beam radiation therapy alone for locally advanced prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2000; 2000/08/02:2869-2880.
- (117) Wilder RB, Barme GA, Gilbert RF, Holevas RE, Kobashi LI, Reed RR et al. Preliminary results in prostate cancer patients treated with high-dose-rate brachytherapy and intensity modulated radiation therapy (IMRT) vs. IMRT alone. *Brachytherapy* 2010; 2009/10/27:341-348.
- (118) Soumarova R, Homola L, Perkova H, Stursa M. Three-dimensional conformal external beam radiotherapy versus the combination of external radiotherapy with high-dose rate brachytherapy in localized carcinoma of the prostate: comparison of acute toxicity. *Tumori* 2007; 2007/04/26:37-44.
- (119) Joseph KJ, Alvi R, Skarsgard D, Tonita J, Pervez N, Small C et al. Analysis of health related quality of life (HRQoL) of patients with clinically localized prostate cancer, one year after treatment with external beam radiotherapy (EBRT) alone versus EBRT and high dose rate brachytherapy (HDRBT). *Radiation oncology (London, England)* 2008; 2008/07/17:20.
- (120) Sylvester JE, Grimm PD, Blasko JC, Millar J, Orio PF, Skoglund S et al. 15-Year biochemical relapse free survival in clinical Stage T1-T3 prostate cancer following combined external beam radiotherapy and brachytherapy; Seattle experience. *International Journal of Radiation Oncology* Biology* Physics* 2007; 67(1):57-64.
- (121) Klotz L, Boccon-Gibod L, Shore ND, Andreou C, Persson BE, Cantor P et al. The efficacy and safety of degarelix: a 12-month, comparative, randomized, open-label, parallel-group phase III study in patients with prostate cancer. *BJU international* 2008; 102(11):1531-1538.
- (122) Crook JM, O'Callaghan CJ, Duncan G, Dearnaley DP, Higano CS, Horwitz EM et al. Intermittent androgen suppression for rising PSA level after radiotherapy. *The New England journal of medicine* 2012; 2012/08/31:895-903.
- (123) Gerger A, Langsenlehner U, Kapp K, Langsenlehner T. Overall survival benefit and safety profile of radium-223 chloride, a first-in-class alpha-pharmaceutical: results from a phase III randomized trial (ALSYMPCA) in patients with castration-resistant prostate cancer (CRPC) with bone metastases. *Journal of Clinical Oncology* 2012; 30(Suppl 5).
- (124) Vora SA, Wong WW, Schild SE, Ezzell GA, Halyard MY. Analysis of biochemical control and prognostic factors in patients treated with either low-dose three-dimensional conformal radiation therapy or high-dose intensity-modulated

radiotherapy for localized prostate cancer. *International journal of radiation oncology, biology, physics* 2007; 2007/04/03:1053-1058.

- (125) National Institute for Health and Clinical E. Atrial fibrillation: the management of atrial fibrillation: NICE clinical guideline 36. United Kingdom: National Institute for Health and Clinical Excellence (NICE); 2006.
- (126) Camm AJ, Capucci A, Hohnloser SH, Torp-Pedersen C, Van Gelder IC, Mangal B et al. A randomized active-controlled study comparing the efficacy and safety of vernakalant to amiodarone in recent-onset atrial fibrillation. *J Am Coll Cardiol* 2011; 57(3):313-321.
- (127) National Collaborating Centre for Chronic Conditions. Atrial fibrillation: National clinical guideline for management in primary and secondary care. London: Royal College of Physicians; 2006.
- (128) National Institute for Health and Clinical E. Final scope - Atrial fibrillation: The management of atrial fibrillation. United Kingdom: National Institute for Health and Clinical Excellence (NICE); 2004.
- (129) Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S et al. 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. *Health Technol Assess* 2005; 9(40):1-90.
- (130) Bayer Healthcare. Single Technology Appraisal (STA) of rivaroxaban (Xarelto®) for the prevention of venous thromboembolism (VTE) in adult patients undergoing elective hip or knee replacement surgery. 2009.
- (131) Boehringer-Ingelheim. Venous thromboembolism - dabigatran: manufacturer submission. 2008.
- (132) Jowett S, Bryan S, Mant J, Fletcher K, Roalfe A, Fitzmaurice D et al. Cost Effectiveness of Warfarin Versus Aspirin in Patients Older Than 75 Years With Atrial Fibrillation. *Stroke* 2011; 42(6):1717-1721.
- (133) Kansal AR, Sorensen SV, Gani R, Robinson P, Pan F, Plumb JM et al. Cost-effectiveness of dabigatran etexilate for the prevention of stroke and systemic embolism in UK patients with atrial fibrillation. *Heart* 2012; 98(7):573-578.
- (134) McKeage K. Dabigatran Etexilate: A Pharmacoeconomic Review of its Use in the Prevention of Stroke and Systemic Embolism in Patients with Atrial Fibrillation. *Pharmacoeconomics* 2012; 30(9):841-855.
- (135) Sullivan PW, Arant TW, Ellis SL, Ulrich H. The Cost Effectiveness of Anticoagulation Management Services for Patients with Atrial Fibrillation and at High Risk of Stroke in the US. *Pharmacoeconomics* 2006; 24(10):1021-1033.

- (136) Edwards SJ, Hamilton V, Nherera L, Trevor N, Barton S. Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation. A Single Technology Appraisal. London: BMJ-TAG; 2011.
- (137) National Institute for Health and Clinical Excellence. Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation. NICE technology appraisal guidance 249. London: NICE; 2012.
- (138) National Institute for Health and Clinical Excellence. Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation. NICE technology appraisal guidance TA256. London: NICE; 2012.
- (139) Spackman E, Burch J, Faria R, Corbacho B, Fox D, Woolacott N. Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation. A Single Technology Appraisal. University of York: Centre for Reviews and Dissemination and Centre for Health Economics; 2011.
- (140) McKenna C, Maund E, Sarowar M, Fox D, Stevenson M, Pepper C et al. Dronedronone for atrial fibrillation and flutter A Single Technology Appraisal. Evidence Review Group Report. Centre for Reviews and Dissemination (CRD), Centre for Health Economics (CHE); 2009.
- (141) National Institute for Health and Clinical Excellence. Dronedronone for the treatment of non-permanent atrial fibrillation: Technology appraisal TA197. London: NICE; 2010.
- (142) Sanofi Aventis. Single technology appraisal of Multaq (Dronedronone). Manufacturer/sponsor submission of evidence. 2009.
- (143) Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S et al. Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation. 2008; 12(34):1-220.
- (144) Andrikopoulos G, Tzeis S, Maniadakis N, Mavrakis HE, Vardas PE. Cost-effectiveness of atrial fibrillation catheter ablation. 2009; 11(2):147-151.
- (145) McKenna C, Palmer S, Rodgers M, Chambers D, Hawkins N, Golder S et al. Cost-effectiveness of radiofrequency catheter ablation for the treatment of atrial fibrillation in the United Kingdom. 2009; 95(7):542-549.
- (146) Quenneville SP, Xie X, Brophy JM. The cost-effectiveness of Maze procedures using ablation techniques at the time of mitral valve surgery. 2009; 25(4):485-496.
- (147) Reynolds MR, Zimetbaum P, Josephson ME, Ellis E, Danilov T, Cohen DJ. Cost-effectiveness of radiofrequency catheter ablation compared with antiarrhythmic drug therapy for paroxysmal atrial fibrillation. 2009; 2(4):362-369.
- (148) Chen HS, Wen JM, Wu SN, Liu JP. Catheter ablation for paroxysmal and persistent atrial fibrillation. *Cochrane Database Syst Rev* 2012; 4.

- (149) Friberg L, Rosenqvist M, Lip GYH. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. *Eur Heart J* 2012; 33(12):1500-1510.
- (150) Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach The Euro Heart Survey on Atrial Fibrillation. *CHEST Journal* 2010; 137(2):263-272.
- (151) Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A Novel User-Friendly Score (HAS-BLED) To Assess 1-Year Risk of Major Bleeding in Patients With Atrial Fibrillation The Euro Heart Survey. *CHEST Journal* 2010; 138(5):1093-1100.
- (152) Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991; 121(1):293-298.
- (153) Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Ann Intern Med* 2008; 148(2):102.
- (154) Wilson PF MJS. Prediction of incident diabetes mellitus in middle-aged adults: The framingham offspring study. *Arch Intern Med* 2007; 167(10):1068-1074.
- (155) Cowie MR, Wood DA, Coats AJS, Thompson SG, Poole-Wilson PA, Suresh V et al. Incidence and aetiology of heart failure; a population-based study. *Eur Heart J* 1999; 20(6):421-428.
- (156) Hughes M, Lip GYH. Stroke and thromboembolism in atrial fibrillation: a systematic review of stroke risk factors, risk stratification schema and cost effectiveness data. 2008; 99(2):295-304.
- (157) Lip GYH, Frison L, Halperin JL, Lane DA. Identifying Patients at High Risk for Stroke Despite Anticoagulation A Comparison of Contemporary Stroke Risk Stratification Schemes in an Anticoagulated Atrial Fibrillation Cohort. *Stroke* 2010; 41(12):2731-2738.
- (158) Lip GYH, Halperin JL. Improving Stroke Risk Stratification in Atrial Fibrillation. *The American Journal of Medicine* 2010; 123(6):484-488.
- (159) Hobbs DR, Roalfe AK, Lip YH, Fletcher K, Fitzmaurice DA, Mant J. Performance of stroke risk scores in older people with atrial fibrillation not taking warfarin: comparative cohort study from BAFTA trial. *BMJ* 2011; 342:1-13.
- (160) Olesen JB, Lip GYH, Hansen ML, Hansen PR, Tolstrup JS, Lindhardsen J et al. Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. *BMJ: British Medical Journal* 2011; 342.

- (161) Van Staa TP, Setakis E, Di Tanna GL, Lane DA, Lip GYH. A comparison of risk stratification schemes for stroke in 79 884 atrial fibrillation patients in general practice. *Journal of Thrombosis and Haemostasis* 2011; 9(1):39-48.
- (162) Gage BF WAS. Validation of clinical classification schemes for predicting stroke: Results from the national registry of atrial fibrillation. *JAMA: The Journal of the American Medical Association* 2001; 285(22):2864-2870.
- (163) Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. *Arch Intern Med* 1994; 154:1449-1457.
- (164) Lip GYH, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) score. *J Am Coll Cardiol* 2011; 57(2):173-180.
- (165) Lip GYH, Andreotti F, Fauchier L, Huber K, Hylek E, Knight E et al. Bleeding risk assessment and management in atrial fibrillation patients: a position document from the European Heart Rhythm Association, endorsed by the European Society of Cardiology Working Group on Thrombosis. *Europace* 2011; 13(5):723-746.
- (166) Nieuwlaat R, Prins MH, Le Heuzey JY, Vardas PE, Aliot E, Santini M et al. Prognosis, disease progression, and treatment of atrial fibrillation patients during 1 year: follow-up of the Euro Heart Survey on atrial fibrillation. *Eur Heart J* 2008; 29(9):1181-1189.
- (167) Kerr CR, Humphries KH, Talajic M, Klein GJ, Connolly SJ, Green M et al. Progression to chronic atrial fibrillation after the initial diagnosis of paroxysmal atrial fibrillation: results from the Canadian Registry of Atrial Fibrillation. *Am Heart J* 2005; 149(3):489.
- (168) Reiffel JA, Schwarzberg R, Murry M. Comparison of autotriggered memory loop recorders versus standard loop recorders versus 24-hour holter monitors for arrhythmia detection. *Am J Cardiol* 2005; 95(9):1055-1059.
- (169) Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W et al. Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation. *N Engl J Med* 2011; 365(10):883-891.
- (170) Freemantle N, Lafuente-Lafuente C, Mitchell S, Eckert L, Reynolds M. Mixed treatment comparison of dronedarone, amiodarone, sotalolol, flecainide, and propafenone, for the management of atrial fibrillation. *Europace* 2011; 13(3):329-345.
- (171) Lang R, Klein HO, Weiss E. Superiority of oral verapamil therapy to digoxin in treatment of chronic atrial fibrillation. *Chest* 1983; 83(3):491-499.

- (172) Lewis RV, Laing E, Moreland TA, Service E, McDevitt DG. A comparison of digoxin, diltiazem and their combination in the treatment of atrial fibrillation. *Eur Heart J* 1988; 9(3):279-283.
- (173) Roth A, Harrison E, Mitani G. Efficacy and safety of medium- and high-dose diltiazem alone and in combination with digoxin for control of heart-rate at rest and during exercise in patients with chronic atrial fibrillation. *Circulation* 1986; 73(2):316-324.
- (174) Wong C-K, Lau C-P, Leung W-H, Cheng CH. Usefulness of labetalol in chronic atrial fibrillation. *Am J Cardiol* 1990; 66(17):1212-1215.
- (175) Stewart S, MacIntyre K, MacLeod MMC, Bailey AEM, Capewell S, McMurray JJV. Trends in hospital activity, morbidity and case fatality related to atrial fibrillation in Scotland, 1986-1996. *Eur Heart J* 2001; 22(8):693-701.
- (176) Luengo-Fernandez R, Yiin GSC, Gray AM, Rothwell PM. Population-based study of acute- and long-term care costs after stroke in patients with AF. *International Journal of Stroke* 2012;1-7.
- (177) Douketis JD, Arneklev K, Goldhaber SZ, Spandorfer J, Halperin F, Horrow J. Comparison of bleeding in patients with nonvalvular atrial fibrillation treated with ximelagatran or warfarin: assessment of incidence, case-fatality rate, time course and sites of bleeding, and risk factors for bleeding. *Arch Intern Med* 2006; 166(8):853.
- (178) Ara R, Wailoo A. Using Health State Utility Values in Models Exploring the Cost-Effectiveness of Health Technologies. *Value in Health* 2012; 15(6):971-974.
- (179) Steg PG, Alam S, Chiang CE, Gamra H, Goethals M, Inoue H et al. Symptoms, functional status and quality of life in patients with controlled and uncontrolled atrial fibrillation: data from the RealiseAF cross-sectional international registry. *Heart* 2012; 98(3):195-201.
- (180) Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. *Medical care* 2005; 43(7):736-749.
- (181) Chevalier P, Durand-Dubief A, Burri H, Cucherat M, Kirkorian G, Touboul P. Amiodarone versus placebo and class IC drugs for cardioversion of recent-onset atrial fibrillation: A meta-analysis. *J Am Coll Cardiol* 2003; 41(2):255-262.
- (182) Bash LD, Buono JL, Davies GM, Martin A, Fahrback K, Phatak H et al. Systematic review and meta-analysis of the efficacy of cardioversion by vernakalant and comparators in patients with atrial fibrillation. *Cardiovascular drugs and therapy* 2012; 26:167-179.
- (183) Olshansky B, Rosenfeld LE, Warner AL, Solomon AJ, O'Neill G, Sharma A et al. The Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study:

- approaches to control rate in atrial fibrillation. *J Am Coll Cardiol* 2004; 43(7):1201-1208.
- (184) Wyse DG, Waldo AL, DiMarco JP, Domanski MJ, Rosenberg Y, Schron EB et al. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med* 2002; 347(23):1825-1833.
- (185) Corley SD, Epstein AE, DiMarco JP, Domanski MJ, Geller N, Greene HL et al. Relationships between sinus rhythm, treatment, and survival in the Atrial Fibrillation Follow-Up Investigation of Rhythm Management (AFFIRM) Study. *Circulation* 2004; 109(12):1509.
- (186) National Institute for Health and Clinical E. Review consultation document. Review of Clinical Guideline CG36 Atrial Fibrillation. United Kingdom: National Institute for Health and Clinical Excellence (NICE); 2011.
- (187) De Paola AA, Figueiredo E, Sesso R, Veloso HH, Nascimento LO. Effectiveness and costs of chemical versus electrical cardioversion of atrial fibrillation. *Int J Cardiol* 2003; 88(2-3):157-166.
- (188) Cordina J, Mead G. Pharmacological cardioversion for atrial fibrillation and flutter (Review). *The Cochrane Library* 2009; 1.
- (189) Mead G, Elder A, Flapan AD, Cordina J. Electrical cardioversion for atrial fibrillation and flutter (Review). *The Cochrane Library* 2009; 1.
- (190) Rienstra M, Van Veldhuisen DJ, Crijns HJGM, Van Gelder IC. Enhanced cardiovascular morbidity and mortality during rhythm control treatment in persistent atrial fibrillation in hypertensives: data of the RACE study. *Eur Heart J* 2007; 28(6):741-751.
- (191) Roy D, Talajic M, Nattel S, Wyse DG, Dorian P, Lee KL et al. Rhythm control versus rate control for atrial fibrillation and heart failure. *N Engl J Med* 2008; 358(25):2667-2677.
- (192) Nilsson Jr KR, Al-Khatib SM, Zhou Y, Pieper K, White HD, Maggioni AP et al. Atrial fibrillation management strategies and early mortality after myocardial infarction: results from the Valsartan in Acute Myocardial Infarction (VALIANT) Trial. *Heart* 2010; 96(11):838-842.
- (193) Camm AJ, Kirchhof P, Lip GYH, Schotten U, Savelieva I, Ernst S et al. Guidelines for the management of atrial fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J* 2010; 31(19):2369-2429.
- (194) Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Medical care* 2005; 43(3):221-228.

- (195) National Institute for Health and Clinical Excellence. Diagnostics Assessment Programme manual. London: NICE; 2011.
- (196) Hoyle M, Anderson R. Whose Costs and Benefits? Why Economic Evaluations Should Simulate Both Prevalent and All Future Incident Patient Cohorts. *Med Decis Mak* 2010; 30(4):426-437.
- (197) Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *The European Journal of Health Economics* 2008; 9(3):251-259.

Appendix 1. MAPGuide study protocol

1. Title

Economic modelling of diagnostic/treatment pathways in NICE clinical guidelines: feasibility and value for informing decisions about updates

2. Importance

NICE clinical guidelines provide advice on appropriate diagnosis and care for people with specific diseases and conditions in the NHS in England and Wales [1]. As of July 2009, 88 guidelines had been published, covering a diverse range of patient groups and conditions, and 46 guidelines were in development. Though compliance with NICE guidelines is not compulsory, they set standards for NHS organisations and professionals, and have a major impact on patient care.

Guidelines are developed for NICE by four National Collaborating Centres (NCCs) and an in-house 'short guidelines' team. As with technology appraisals and public health guidance, groups developing NICE guidelines are expected to take account of cost effectiveness [2,3]. Health economists work alongside other NCC technical staff, healthcare professionals and patient representatives in Guideline Development Groups (GDGs). The special role of the economist is to provide evidence on cost-effectiveness and advice on how this should be interpreted. But this is difficult because of the size and complexity of NICE guidelines, which may cover up to 30 questions along a 'pathway' of care (or 'algorithm'). This may include aspects of assessment, diagnosis, treatment, long-term management and follow-up, and, although few guidelines cover the whole pathway and recent efforts aim to produce more focussed guidelines, NICE guidelines remain large and complex pieces of work. Guideline economists cannot evaluate every clinical question in great depth, but instead use a selective approach; relying on published economic evidence when this is of sufficient quality and relevance, conducting new analyses for key questions and encouraging the GDG to use judgement about the broad balance of benefits, harms and costs for the remaining issues [1]. This approach is pragmatic, and may be good enough, ensuring that the really important economic issues are identified and addressed [4].

However, it is also possible that important issues are being missed or inadequately considered. Alan Williams considered this dilemma in his 2004 OHE lecture [5], concluding:

"I think that guideline development needs to be strengthened from the outset by injecting into the process a strong dose of decision-analytic expertise, so as to ensure that the whole territory is mapped out in a systematic way. ... we need not only a large-scale map of what to do at particularly tricky junctions, but also a small-scale map of the entire system covering all the relevant highways and byways, and estimating the traffic flows along each."

A piecemeal approach to economic analysis may ignore important connections and feedback in the patient pathway. For example, the sequencing of tests and treatments may radically alter costs and health outcomes. The cost-effectiveness of a test depends on downstream treatment decisions, and conversely the cost-

effectiveness of a treatment depends on upstream selection of patients. Health economists and decision analysts are very familiar with these complexities [6], which can be integrated by adopting a wider boundary around the model and capturing a greater breadth in the pathway of care. Whether or not NICE guidelines continue to define a formal pathway, it is important that recommendations are evaluated within this broader context. A recent initiative, funded by the Department of Health, has demonstrated this approach by building a 'whole disease' model for colorectal cancer [7]. Discrete event simulation was used to model current practice, following patients from initial presentation through to end-of-life care. This model was then used as a comparator, against which the cost-effectiveness of potential service developments was evaluated. In an NIHR fellowship building on this work, Paul Tappenden is now developing a methods framework for whole disease models of cancer. This fellowship study will provide guidance on the approach and how it may be usefully implemented.

The idea of building a model of the patient pathway to serve as a foundation for economic evaluation in NICE guidelines is attractive. However, it may not be feasible given the challenging deadlines and resource constraints of 'live' guideline development. We therefore propose to test the approach first within the simpler context of guideline update decisions. The NICE guidelines programme has moved into a phase where hard decisions are needed to balance demands to maintain the backlog of published guidelines with demands for new guidelines to address emerging priorities. This creates a logistical problem for NICE, but also an opportunity to test the feasibility and usefulness of pathway modelling. Our experience is that eliciting an agreed pathway at the beginning of guideline development can be difficult – guideline topics are usually identified precisely because there is high uncertainty or disagreement about what is, or should be, standard practice. We will therefore start by selecting case studies of existing guidelines with well-articulated pathways of care, and modelling these recommended pathways.

On their own, the pathway models will be of little use for decision-making; they may help us to estimate the cost impact or burden of disease associated with a set of services, but cannot tell us how cost-effective those services are. For that, we need to compare the standard pathway with some variations, reflecting possible options for change – for example, substituting a different test or treatment at a given point in the pathway. In the current process for updating NICE guidelines, NCCs search for new evidence and suggestions for possible update topics [1]. We will observe this process for our case studies, and collate lists of suggested update topics. We will then adapt the models to estimate the incremental cost-effectiveness of potential changes to the pathway. These estimates will be subject to considerable uncertainty, but by modelling this uncertainty we should be able to estimate the maximum value of updating each topic using a 'value of information' approach [8]. This should help to identify aspects of the pathway that are both sensitive to change and where the change would have an important impact on patient outcomes and NHS costs. Any such topics may represent priorities for update. Or, if no such topics are identified; this may suggest that an update is not warranted.

Finally, we will examine whether the priorities for update suggested by this modelling approach differ from those that would be identified anyway. This will be assessed through a form of 'Delphi' survey, in which

people consulted during the routine updating process will be invited to rate topics first without the results of the modelling, and then again with that information.

In summary, our primary research aims are:

- To investigate the feasibility of modelling pathways recommended in NICE clinical guidelines to estimate associated patient flows, health outcomes and costs.
- To illustrate how such models can be used as a basis for assessing the incremental cost-effectiveness of possible variations in the care pathway.
- To use this approach to estimate the value of updating selected topics within a guideline.
- To compare the update priorities obtained from formal modelling with those elicited from participants in the routine updating process.

In order to achieve these aims, we will:

- A. Select two NICE guidelines to serve as illustrative examples.
- B. For each guideline, build a simulation model of the recommended pathway to estimate overall patient flows, health outcomes and costs if the guideline is followed.
- C. Collate suggestions for update topics and sources of new evidence by observing the NCC-led updating process.
- D. Ask participants in the update process (NCC/NICE staff, clinical experts and patient representatives) to rate the suggested topics in terms of priority for inclusion in an update.
- E. Adapt the models to estimate the incremental net benefit of possible changes to pathways.
- F. Use a value of information approach to estimate the maximum expected net benefit of updating each suggested topic.
- G. Feedback the results from steps E and F to the people consulted in step D, and invite them to reassess their ratings of priorities for update.

In addition to addressing an important methodological issue, this research could also be of practical use to NICE, helping to inform decisions about updating the guidelines that we use for case studies. Building on our experience, we will also make recommendations for future research into the broader potential for pathway modelling in guideline development, and on possible strategies for deploying this approach during routine guideline development.

3. Scientific potential

People and track record

The Principal Investigator, **Joanne Lord**, will take overall responsibility for leadership and management of the project:

- JL is a reader in the Health Economics Research Group (HERG) at Brunel University. She worked for four years in the guidelines team at NICE, where she provided advice within the Institute and to NCCs on the use of economic evaluation in clinical guidelines. Before joining NICE, she was a lecturer in health economics at Imperial College Management School and in the Public Health Department at St George's Hospital Medical School. She has conducted applied economic evaluations based on clinical trials and modelling studies, and published on methodological aspects of economic evaluation.

JL will also lead the Brunel modelling team, which will develop and apply the model for one of the case studies. Other members of this team are: **Julie Eatock**, who will lead on building the model; and **Gethin Griffith**, who will lead on the collection of data to populate the model.

- JE is a research fellow in the Department of Information Systems and Computing (DISC) at Brunel University. Her main research interest is in applying advanced discrete event simulation techniques to better model systems, and hence improve decision analysis. She has applied this modelling technique in many different contexts including: information systems; business processes; new product development for medical devices; evaluation of telemedicine; and the A&E department of a local hospital.
- GG is a research fellow in health economics in HERG. Prior to joining HERG in 2006, he was a research fellow with the Centre for the Economics of Health, University of Wales Bangor (UWB), the Health Services Research Unit (UWB) and has worked for Research Centre Wales (UWB). His research interests include the economic evaluation of genetic health services, decision analytic modelling and discrete choice modelling.

The second modelling team, based at the London School of Hygiene and Tropical Medicine (LSHTM), will be led by **Alec Miners**. **Bernadette Li** and **Sarah Willis** will lead on data collection and modelling respectively.

- AM is a lecturer in health economics in the Health Services Research Unit (HSRU) at the LSHTM. Following graduation from the MSc in Health Economics at York, he worked for five years at the Royal Free Medical School, London. He then went on to work in the Centre for Health Technology Assessment at NICE as a Technical Advisor and as a honorary Research Fellow at Brunel University, before joining the LSHTM in 2006. He is a member of NICE's Technology Appraisals Committee and Decision Support Unit. He leads a team of economists who provide analytical support to the National Collaborating Centre for Cancer (NCC-C).
- SW is a research fellow in the HSRU. She has been working for over two years as a health economist on NICE cancer guidelines. She led the economic programme of work for two guidelines: advanced breast cancer (published 2009) and lung cancer (anticipated 2011). She has developed a large decision tree model to assess the cost-effectiveness of sequences of chemotherapy. This model was informed by an indirect treatment comparison model developed in collaboration with researchers at Bristol University.
- BL joined the HSRU as a research fellow working on NICE cancer guidelines in 2008. She provides health economics input and advice to GDGs, undertakes economic modelling for high priority areas, and supervises two research assistants. She has experience as a Senior Outcomes Research Manager, Health

Outcomes Scientist and Associate Product Manager in the pharmaceutical industry, and as an Editorial Assistant for the editor of an international peer reviewed medical journal.

Simon Taylor will provide external advice to the modelling teams on simulation modelling techniques and application. He will also take responsibility for model verification and validation.

- ST is a reader in DISC. His main research aim is to use novel computing technologies and techniques to benefit operational research. His research interests include distributed systems and computing, simulation modelling, distributed and web-based simulation, applications of Grid computing and the Semantic Web. Before joining Brunel, Simon lectured at Westminster and Leeds Met. University.

Paul Tappenden will join the research team in December 2009, when his NIHR fellowship comes to an end. He will lead on observing the conventional update process and eliciting expert ratings, preventing premature feedback of this information to the modelling teams.

- PT is a senior research fellow in the Health Economics and Decision Science group in the School for Health and Related Research (SchARR) at the University of Sheffield. Since joining SchARR in 2000, he has led or contributed to the modelling work for 8 NICE technology appraisals, and has developed health economic models for the NCCHTA, the Department of Health and NHS Cancer Screening Programmes. He is currently undertaking an NIHR fellowship to develop, implement and evaluate a methodological framework for modelling whole disease areas to inform resource allocation decisions.

In addition to the above researchers, the project steering group includes NCC and NICE economists and reviewers with extensive experience of the NICE guidelines programme:

- **Phil Alderson** is a public health doctor who has worked in the field of research synthesis since 1996 at the UK Cochrane Centre and currently as Associate Director (Methodology) in the guidelines team at NICE. His role there is to oversee the quality control and methodological development of NICE clinical guidelines.
- **Francis Ruiz** is a health economist who joined the guidelines team at NICE as a technical adviser in July 2006, having previously worked in the Institute's Technology Appraisal programme. He provides leadership within the guidelines programme for all aspects of health economic analysis. Prior to joining NICE, Francis worked in clinical data management and health economics in the pharmaceutical sector.
- **Ifigeneia Mavranouzouli** is a senior health economist at the National Collaborating Centre for Mental Health (NCC-MH), based at UCL. She has worked at this centre for over 4 years. Previously, she worked for a year as a health economist at the National Collaborating Centre for Women's and Children's health (NCC-WCH). She has participated in 9 Guideline Development Groups as member of the Technical Team of the NCC-MH and NCC-WCH. She is a medical doctor and has working experience in primary and secondary care settings.
- **Dave Wonderling** is health economics lead at the National Clinical Guidelines Centre (NCGC), based at Royal College of Physicians and Honorary Research Fellow at HERG, Brunel University. He has worked on NICE clinical guidelines since April 2001, participating directly in 9 Guideline Development Groups. He

now co-leads a team of 8 health economists working concurrently on 14 NICE clinical guidelines. He co-authored the first version of the cost-effectiveness chapter of the Guidelines Manual and has been a member of the NICE Clinical Guidelines Joint Methodology Group since its inception. Previously he was a lecturer in health economics at the London School of Hygiene & Tropical Medicine.

- **Paul Jacklin** is a senior health economist at the National Collaborating Centre for Women's and Children's Health (NCC-WCH), and honorary lecturer at the LSHTM. He has worked on 14 clinical and public health guidelines for NICE. Before joining NCC-WCH, he worked at the LSHTM and at Guy's and St Thomas' hospital, where he developed models to evaluate the cost-effectiveness of different diagnostic strategies for coronary artery disease.
- **Maggie Westby** is the clinical effectiveness lead at the National Clinical Guidelines Centre (NCGC), based at Royal College of Physicians. She has worked on NICE clinical guidelines since 2005 and participated in 8 Guideline Development Groups. Before this, she worked in the Royal College of Nursing's guideline programme and before that, for the UK Cochrane Centre on systematic reviewing and its methodology. She currently oversees the clinical effectiveness methodology carried out by 19 systematic reviewers in 14 guidelines.

Environment

All participating academic groups were highly rated in the 2008 Research Assessment Exercise.

The Brunel modelling team brings together staff from two well-established departments with a long, and ongoing, history of effective collaboration. The Health Economics Research Group (HERG) has an international reputation in health economics, developed over more than twenty years. Its focus is on the economic evaluation of a broad range of clinical and health service technologies. HERG is one of six Specialist Research Institutes at Brunel University, which have special status as prestigious centres which have enhanced the University's research base and research income.

The Department of Information Systems and Computing (DISC) is an internationally recognised centre of excellence in biological and healthcare informatics, human-computer interaction, information science, information systems, intelligent data analysis, and software engineering. The Department is home to the largest research group of its type in the country, with many years of experience with simulation modelling, including evaluations of healthcare interventions [9,10].

The LSHTM modelling team currently provide economics input to the NICE National Collaborating Centre for Cancer, working together to develop and deliver complex economic models to tight deadlines. The Health Services Research Unit at the LSHTM was established in 1988, with the aim of carrying out research that helps to improve the quality, organisation and management of health services and systems. Most of their research is in high income countries and, in particular, the UK. Their staff is multi-disciplinary (epidemiology, sociology, psychology, economics, history, statistics, health policy) and multi-professional (nursing, medicine, pharmacy).

Additional expertise, and a degree of external oversight, will be provided by PT, based in the Health Economics and Decision Science (HEDS) group in the School of Health and Related Research (SchARR) at the University of Sheffield. HEDS specialises in the application of economic evaluation and mathematical modelling to the development of health services and the improvement of the public health. The School employs around 200 multidisciplinary staff and attracts in excess of £6 million per year in external support.

The two modelling teams will meet regularly to coordinate activity and exchange information. All researchers and applicants will participate in regular steering group meetings, where the plans and progress of the modelling teams will be presented and discussed.

Senior staff from the National Collaborating Centres (NCCs) with responsibility for the guidelines chosen as case studies will also be invited to participate in steering group meetings as appropriate. This will include early sessions when plans for the scheduling of modelling, observation of the NCC-led update process and expert survey are discussed, as well as later discussions when results are presented and plans for dissemination agreed. This, together with the inclusion of senior technical staff from the guideline programme as co-applicants, will ensure effective liaison with the NCCs and NICE. The steering group will also take responsibility for assuring that the project does not adversely interfere with guideline production by placing an unacceptable burden on NCC staff. Members of the steering group will co-author the case study reports and resulting peer-reviewed publications.

4. Research Plans

A. Select case studies

To allow sufficient time for modelling within the two-year study period, we need guidelines due for update between December 2010 and May 2011 (see project plan below). The steering group will select the case studies when the timetables for update are known, and following consultation with NICE and the NCCs. Other criteria for the selection of the case studies include:

- Existence of a relatively well-formulated pathway in the current guideline.
- Guidelines for different patient groups or disease areas, and which are likely to present different challenges for the modellers.
- Important topics likely to be updated (where the model is likely to have future value)
- But where there is uncertainty/controversy over what topics should be updated

We anticipate that the following guidelines are likely candidates: Prostate Cancer; Irritable Bowel Syndrome; Antenatal Care; Lipid Modification; and Stroke.

B. Model existing pathways

Step 1: Review literature

The modelling teams will start by reviewing literature on published economic models for the disease area and related models from NICE guidance (e.g. technology appraisals). This will help to identify appropriate model structures and sources of data. Documentation for the current NICE guideline will also be reviewed

in detail to ensure understanding of the recommendations, the available evidence and GDG rationale for decisions.

Step 2: Design pathway model

Design model structure based on natural progression of the disease and the recommended guideline pathway. The models will be designed to estimate the number of patients expected to receive different interventions, health outcomes (QALYs) and costs *if* the guideline recommendations were to be fully implemented. Results will be estimated for a population of incident and prevalent cases in England and Wales over the three-year lifetime of the guideline; up to the next point at which the guideline will be reconsidered for update. But to estimate the long-term impact of treatment decisions made during this period, patients entering the pathway will be followed up for life. The models will follow the NICE Reference Case [3]. The model structure and assumptions will be checked with clinical experts and NCC reviewers and economists who are familiar with the guideline.

Step 3: Develop model

The modelling methodology and software will be decided after careful consideration of the requirements for the case studies. A discrete event simulation approach would provide a flexible structure for mapping complicated diagnostic/treatment pathways, and retain information about individual patient history [11,12,13]. Models will be developed following a “rapid prototyping” approach in close collaboration with relevant stakeholders to capture appropriate breadth and depth of detail. The visual animation available in simulation software would also be beneficial in providing a user-friendly interface, enabling better communication with non-economists. This would be important for the methods to be transferable for later use in routine guideline development. However, it must be acknowledged that there can be a technical difficulty with the use of discrete event simulation for economic evaluation, as these models can be very slow to run when combined with probabilistic sensitivity analysis. For example, it did not prove feasible to conduct a value of information analysis of the colorectal ‘whole disease’ model, because of its long run-time [7]. ST and JE will provide advice on the efficient design and use of the models to determine how this problem can be overcome. They will also consider whether and how the modelling approach could be adapted to create a prototype generic modelling tool for pathway analysis in clinical guidelines.

Step 4: Obtain data

Model parameters (incidence and prevalence, baseline risks, test accuracy, treatment effects, utilities and costs) will be fitted using information available in the original guideline, supplemented with new evidence identified through rapid literature searches and/or expert opinion (e.g. by contacting members of the original guideline group). We will not conduct systematic reviews for all of these parameters, as this would not be possible during routine updating. The extent and impact of uncertainty over model parameters will be reflected through probabilistic sensitivity analysis [8].

Step 5: Verification and validation

The modelling teams will check for errors and inconsistencies throughout model development, following best practice for quality assuring simulation [12,13] and decision analytic [14,15] models. The models will be verified internally (to ensure correct programming) and validated externally (to ensure consistency with expected results – for example, that survival times and levels of service use are realistic). In addition, an experienced modeller external to the modelling teams (ST) will independently review the models, and work with the teams to ensure that any identified errors or inconsistencies are corrected.

C. Identify suggestions for update topics

The normal updating process is described in Chapter 14 of the NICE Guidelines Manual [1]. This is led by the NCC, which undertakes literature searches for new evidence and seeks the views of experts, which may include patient representatives and healthcare professionals (often members of the original guideline group). The NCC then makes recommendations to NICE, who decide whether the guideline should be updated, and if so, what are the key areas that need to be considered. This normal updating process will be observed by a researcher working independently from the modelling teams (PT). This may involve attending relevant meetings and reviewing documents as advised by the NCC. From these sources, a list of potential topics for inclusion in an update and a list of any new evidence will be collated. These lists will be supplied to the modelling teams to inform step E and F.

D. Obtain experts' initial ratings of topics

PT will then contact people involved in the update process, which may include NCC/NICE staff as well as any patient representatives and clinicians who were consulted. They will be provided with the list of potential topics identified in step C and asked to rate them in terms of priority for update. PT will liaise with the NCC to identify an appropriate list of people to include in this rating exercise, and to agree procedures for contacting them. The results of the rating exercise will not be given to the modelling teams until after step G.

E. Estimate net benefit of update topics

The modelling teams will identify possible variations in the pathways, which may include:

- substitution of different tests or treatments at given points in the pathway;
- changes to patient eligibility criteria or thresholds for tests or treatments;
- different sequencing of tests or treatments and/or
- addition of tests or treatments as an extra step in the pathway.

Each variation will be evaluated in comparison with the original pathway, to estimate the incremental net benefit of the change for the relevant population (incident and prevalent cases in England and Wales over the three-year lifetime of the guideline). Additional data required to derive these estimates will be obtained from the original guideline, new evidence identified in step C, or by elicitation from experts. Probabilistic sensitivity analysis will be used to estimate the extent of uncertainty over the net benefit estimates. Topics with a greater net benefit offer more potential for gain from a change in recommendations, and are thus a higher priority for inclusion in an update. All other things being equal, net benefits will be greater for topics that affect a large number of patients, offer a large health gain per patient and/or a small increase in costs.

F. Estimate value of information for topics

The modelling teams will then use the cost-effectiveness models to estimate the population Expected Value of Partial Perfect Information (EVPPI) [16] for sets of parameters related to each topic. This provides an estimate of the maximum amount that it would be worth paying to eliminate all uncertainty over the relevant set of parameters. Topics with a larger EVPPI are predicted to offer a greater *potential* for gain if uncertainty over them could be reduced through a guideline update. However, this does not mean that that gain will necessarily be realised, since the update can only summarise the available information – through systematic review of the available literature and further cost-effectiveness modelling. So, in the absence of the necessary primary data, EVPPI will be of limited use in guiding update decisions.

G. Report results to experts and invite revised ratings

The modelling teams will each prepare a report summarising their methods and results. The people who participated in step D will be sent a copy of the modelling report, and invited to comment on it. They will also be presented with their previous ratings of the importance of updating each topic (from step D), alongside estimates of the net benefit and EVPPI for each topic obtained from the modelling (steps E and F). They will then be invited to re-rate the priority of the topics, and to comment on the reasons for their ratings.

5. Ethics and research governance

This study is based on secondary analysis of published data and will not involve any patient contact or use of any individual patient data. We will be contacting experts consulted in the NICE updating process, and will therefore ensure that appropriate ethical approval and research governance are obtained through the Brunel research ethics committee, and LREC if necessary.

6. Data preservation for sharing

No primary data collection. We will provide on-line access to the models developed for this study.

7. Public engagement in science

If the research is successful, the models may be used in subsequent scoping or development of the clinical guidelines. This could involve presentation to participants at stakeholder consultation meetings, and to lay members of guideline development groups.

8. Exploitation and dissemination

For each case study, we will write a final report for the relevant NCC and NICE, summarising the findings from our modelling and the expert surveys. We will also offer NCCs a working copy of the pathway model, along with specific advice and support on its use.

We will present and discuss our overall findings at suitable meetings for NICE and NCC staff (e.g. the Health Economist's in Guidelines meeting, the NCC/NICE technical meeting and the NICE Technical Forum).

We will also disseminate our findings through traditional academic channels, including national and international scientific conferences (e.g. Guidelines International Network, Health Technology Assessment

International, Health Economists' Study Group), and peer-reviewed journals (e.g. Health Economics, Journal of Health Economics, Medical Decision Making, Health Care Technology Assessment).

	Months	Work days						2010		2011	
		JL	JE	GG	ST	PT	Total				
PLANNING											
Select topics, liaise with NCCs and NICE re timetable											
Steering group meetings											
MODEL 1 - Brunel		JL	JE	GG	ST	PT	Total				
1 Review literature	2	4	8	8			20				
2 Design model	1	4	8	8	4		24				
3 Build model	4	4	32	4			40				
4 Collect data	4	4		32			36				
5 Validate model	1				8		8				
6 Revise model as necessary following validation	1	2	4	4			10				
7 Collect update topics by observing NCC/NICE process	2					8	8				
8 Survey experts for priorities (round 1 Delphi)	1				4	4	4				
9 Adapt model to evaluate update suggestions	3	3	20				23				
10 Collect data to test update suggestions	3	3		20			23				
11 Run model and write-up	1	4	8	8			20				
12 Feedback results to experts (round 2 Delphi)	1						0				
13 Write up case study and report to NICE/NCCs	1	4	8	8		4	24				
<i>total:</i>		32	88	88	16	16	240				
MODEL 2 - LSHTM		AM	SW	BL	ST	PT	Total				
14 Review literature	2	4	8	8			20				
15 Design model	1	4	8	8	4		24				
16 Build model	4	4	32	4			40				
17 Collect data	4	4		32			36				
18 Validate model	1				8		8				
19 Revise model as necessary following validation	1	2	4	4			10				
20 Collect update topics by observing NCC/NICE process	2					8	8				
21 Survey experts for priorities (round 1 Delphi)	1				4	4	4				
22 Adapt model to evaluate update suggestions	3	3	20				23				
23 Collect data to test update suggestions	3	3		20			23				
24 Run model and write-up	1	4	8	8			20				
25 Feedback results to experts (round 2 Delphi)	1						0				
26 Write up case study and report to NICE/NCCs	1	4	8	8		4	24				
<i>total:</i>		32	88	88	16	16	240				

- Steering group
- Brunel modelling team (JL, GG, JE)
- LSHTM modelling team (AM, BL, SW)
- External (PT, ST)
- X Guideline update (NCC)

Appendix 2. Guidelines considered as case studies

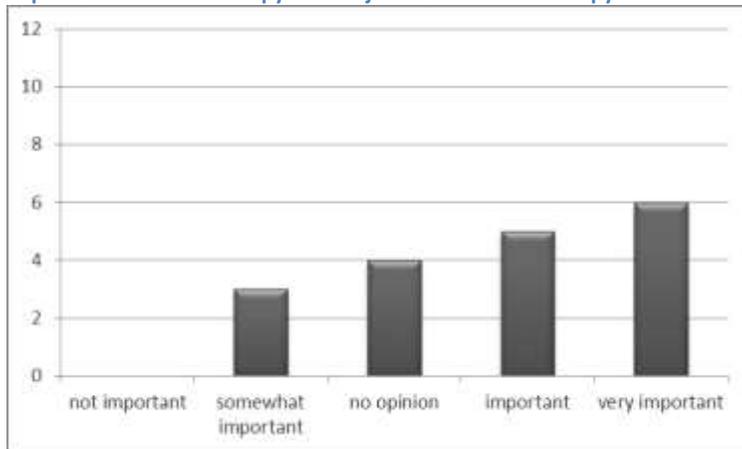
Ref	Title	Issued	Update	Developer	Update NCC	Pathway	Evidence	Update likely	Other comments
CG58	Prostate cancer	Feb-08	Feb-11	NCCC	NCCC	Good	Good	Probably	Is this too similar to PT's colorectal model? Or is that an advantage?
CG59	Osteoarthritis	Feb-08	Feb-11	NCCNSC	NCGC	Poor	Mixed	?	Interventions not sequenced
CG61	Irritable bowel syndrome	Feb-08	Feb-11	NCCWCH	NCGC	Fair	Poor	Possibly not	Pathway quite simple, and little data to support anything more complicated.
CG62	Antenatal care	Mar-08	Mar-11	NCCWCH	NCCWCH	None	?	?	Lots of parallel decisions - eg for different screening tests. No clear pathway.
CG63	Diabetes in pregnancy	Mar-08	Mar-11	NCCWCH	NCCWCH	Good	Fair	?	Extensive modelling already done. Could we add to this?
CG65	Perioperative hypothermia (inadvertent)	Apr-08	Apr-11	NCCNSC	NCGC	Fair	?	?	Pathway over short timeframe.
CG66	Diabetes - type 2 (update)	May-08	May-11	NCCCC	NCGC	Good	Good	?	Complicated interface with short guideline on newer agents.
CG67	Lipid modification	May-08	May-11	NCCPC	NCGC	Good	Good	?	Fair amount of modelling done, but could integrate this. Politics difficult?
CG35	Parkinson's disease	Jun-06	Jun-11	NCCCC	NCGC	Poor	?	?	Interventions not sequenced
CG36	Atrial fibrillation	Jun-06	Jun-11	NCCCC	NCGC	Good	?	?	

Ref	Title	Issued	Update	Developer	Update NCC	Pathway	Evidence	Update likely	Other comments
CG38	Bipolar disorder	Jul-06	Jul-11	NCCMH	NCCMH	Fair	Poor?	?	Pathway is complicated - e.g. many medications.
CG68	Stroke	Jul-08	Jul-11	NCCCC	NGGC	Poor	?	?	Only covers initial diagnosis and management. Short pathway.
CG70	Induction of labour	Jul-08	Jul-11	NCCWCH	NCCWCH	Poor	?	?	Short pathway
CG17	Dyspepsia	Aug-04	Aug-11	Newcastle	NGGC	Good	Good	?	Early guideline, developed by Newcastle. Not representative of current processes.
CG71	Familial hypercholesterolaemia	Aug-08	Aug-11	NCCPC	NGGC	Fair	?	?	
CG72	Attention deficit hyperactivity disorder (ADHD)	Sep-08	Sep-11	NCCMH	NCCMH	Fair	Poor	?	
CG73	Chronic kidney disease	Sep-08	Sep-11	NCCCC	NGGC	Good	Good?	?	Does not cover endstage, though this is included in the model.

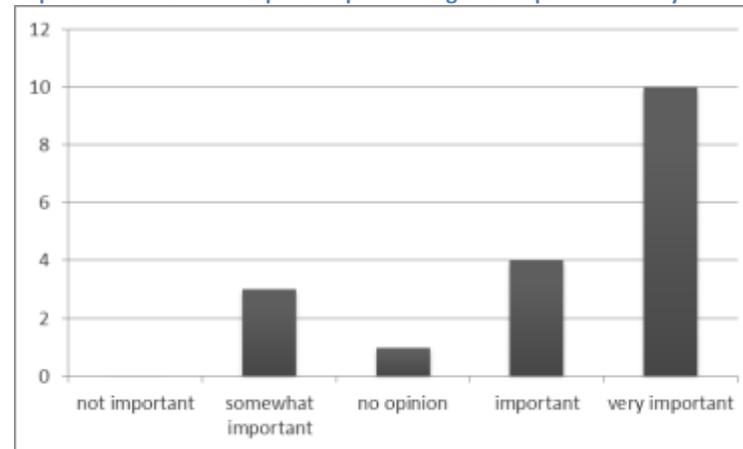
Appendix 3. Stakeholder survey responses

Prostate Cancer survey- Question 1 (Rating of topics) n=18

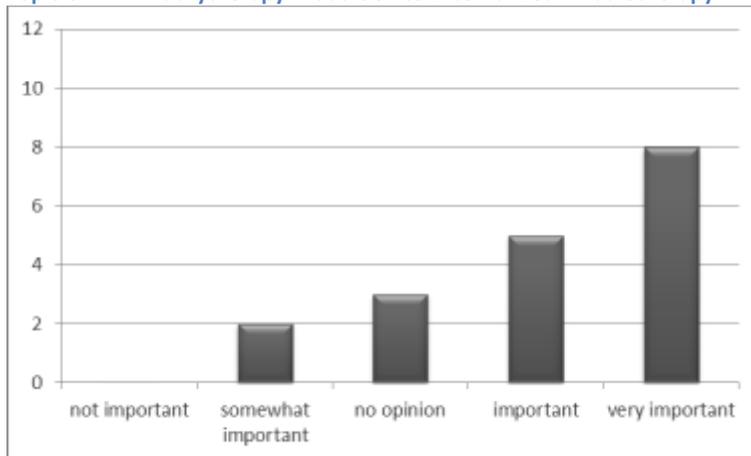
Topic A- Pelvic radiotherapy with adjuvant hormonal therapy



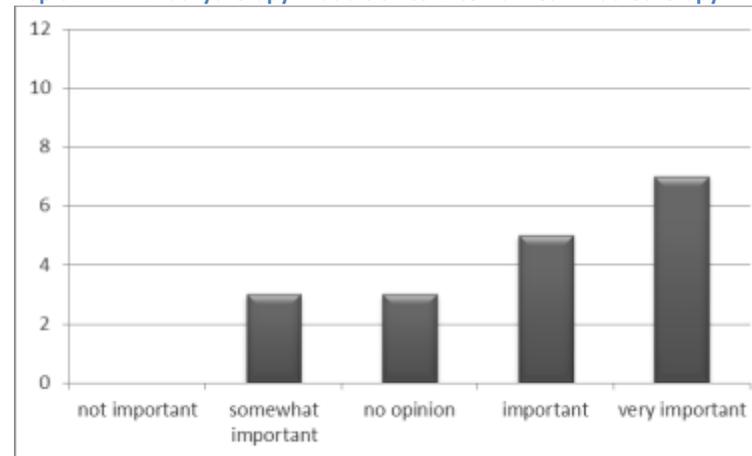
Topic B - Effective techniques for performing radical prostatectomy



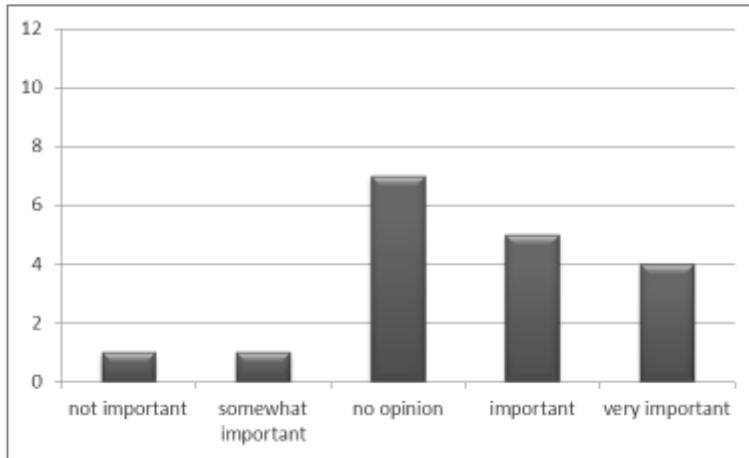
Topic C- HDR Brachytherapy in addition to External Beam Radiotherapy



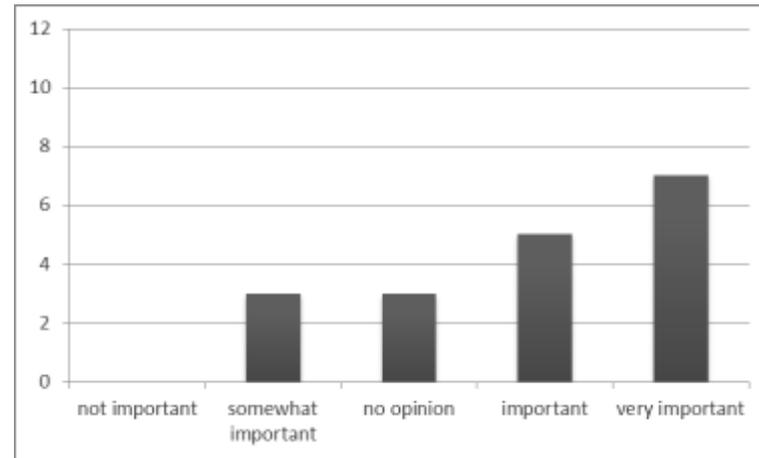
Topic D- LDR Brachytherapy in addition to External Beam Radiotherapy



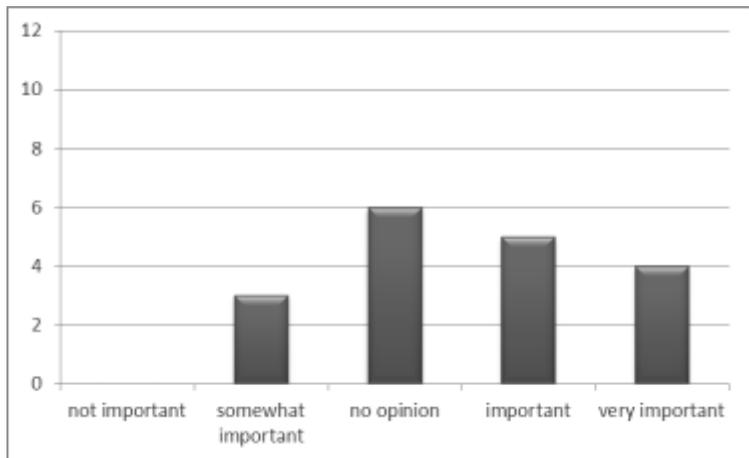
Topic E- Degarelix (a LHRH antagonist) for men with advanced hormone dependent prostate cancer



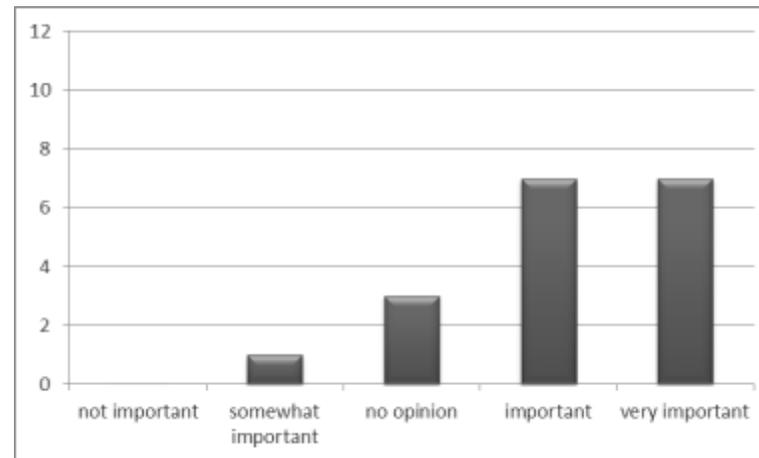
Topic F- Intermittent hormone therapy versus continuous hormone therapy



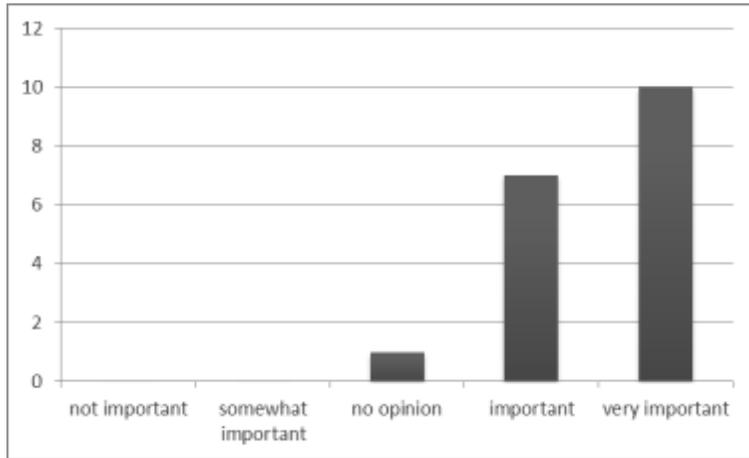
Topic G- Radium 223 chloride versus strontium-89 for men with hormone refractory prostate cancer and painful bone metastases



Topic H- IMRT and IGRT as an alternative to conventional radiotherapy

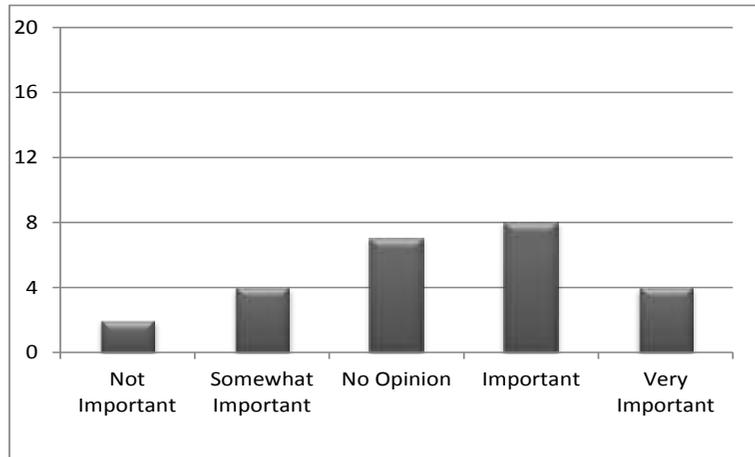


Topic I- Active surveillance in previously unscreened 'low risk' men

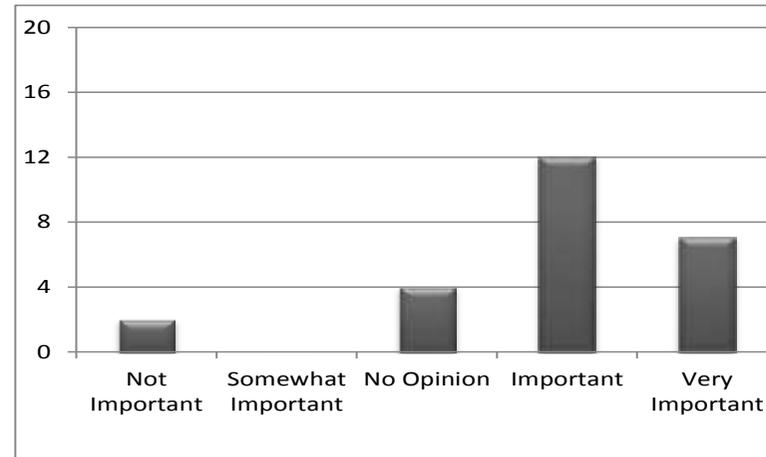


Atrial Fibrillation- Question 1 (Rating of topics) n=25

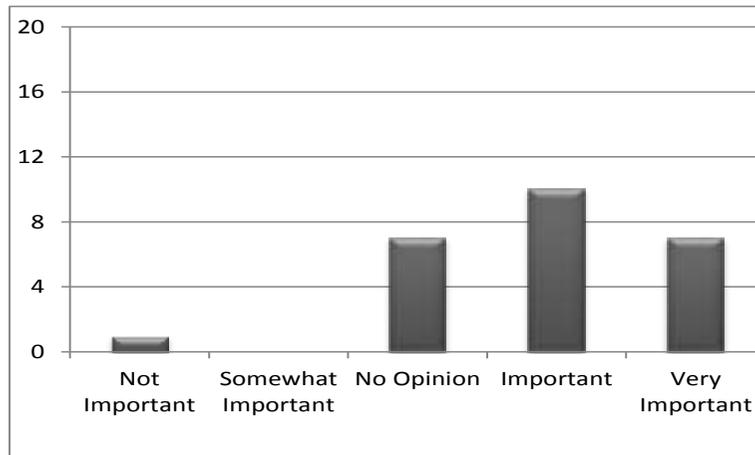
Topic A - Prophylaxis for the prevention of post-operative AF



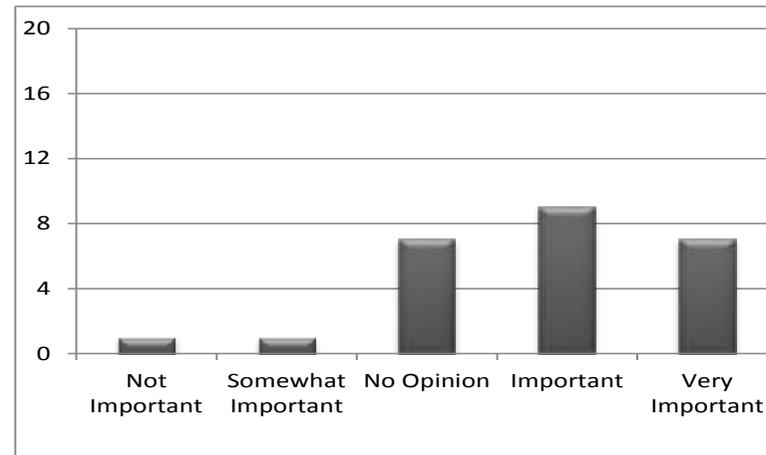
Topic B -Anti-arrhythmic drugs as pharmacological Cardioversion



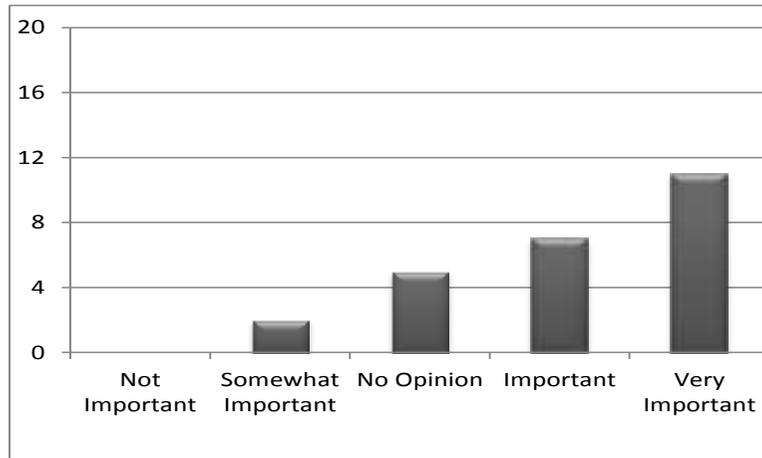
Topic C- Rhythm versus rate control strategies for persistent AF



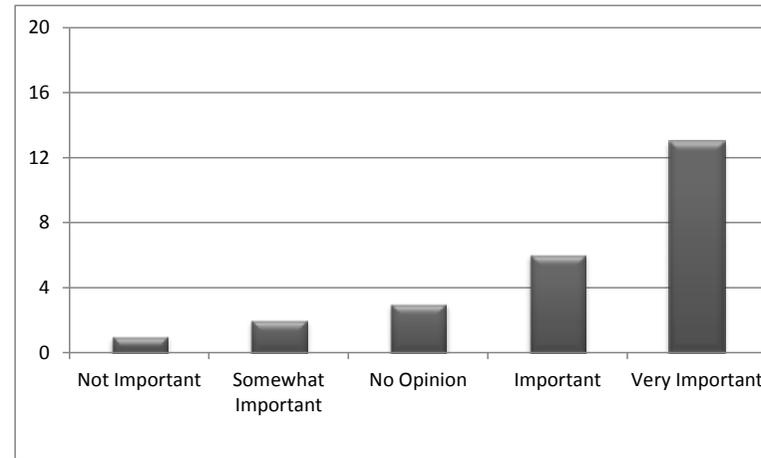
Topic D-Treatment for maintaining Sinus Rhythm after Cardioversion



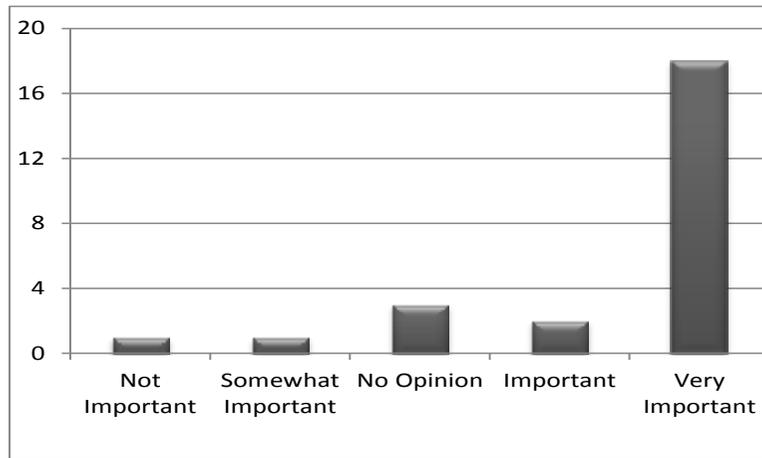
Topic E-Alternative Risk factor based scoring systems to assess stroke risk



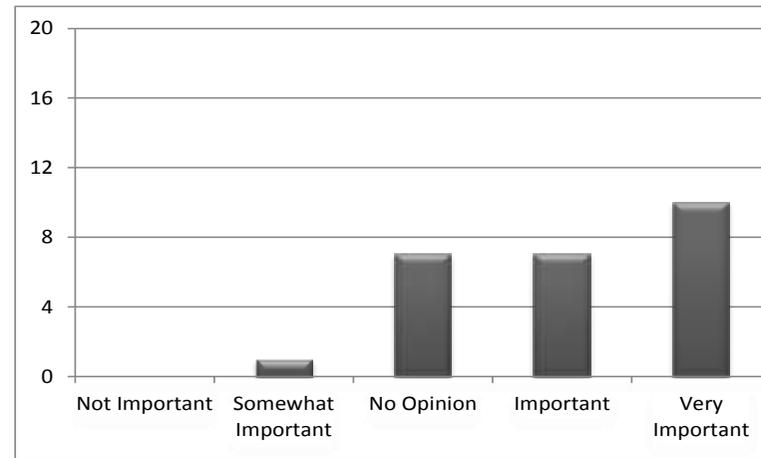
Topic F- Stratification tools to assess bleeding risk



Topic G - Apixaban, Rivaroxaban or Dabigatran etexilate

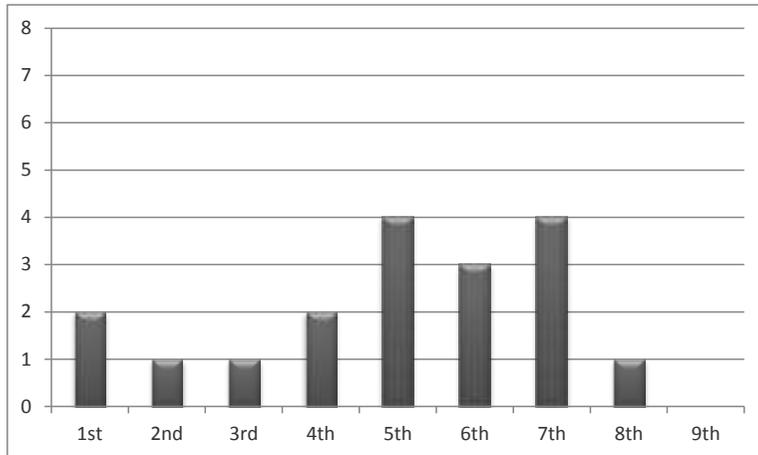


Topic H- Catheter ablation

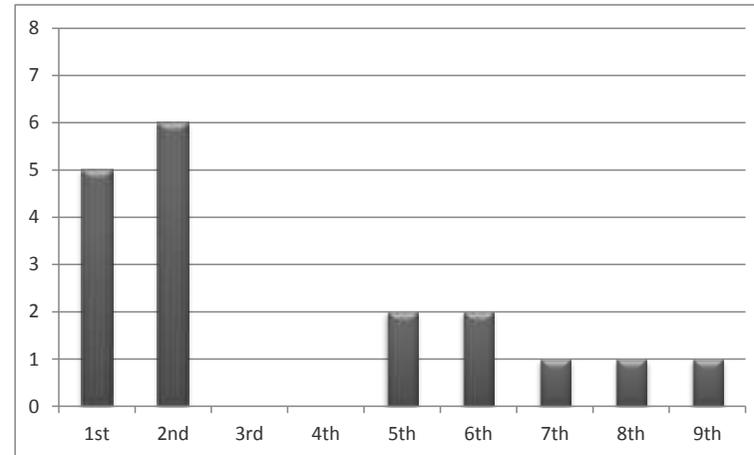


Prostate Cancer- Question 2 (Ranking of topics) n=18

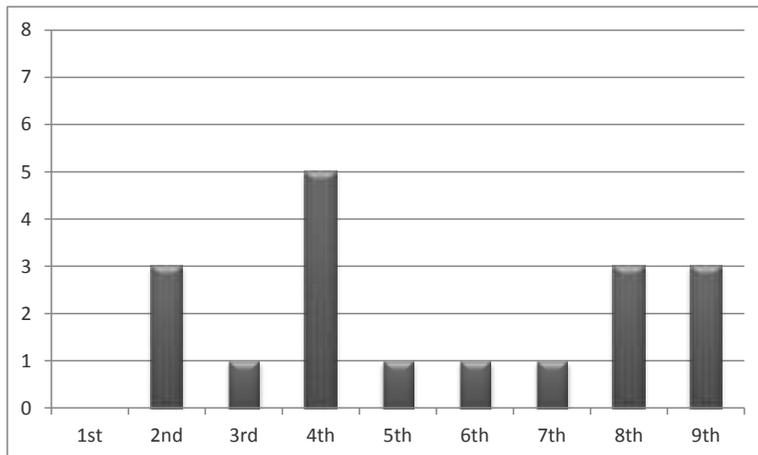
Topic A- Pelvic radiotherapy with adjuvant hormonal therapy



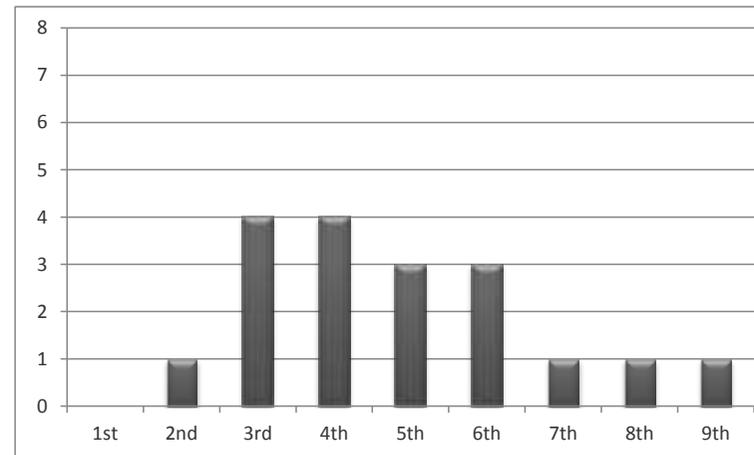
Topic B- Effective techniques for performing radical prostatectomy



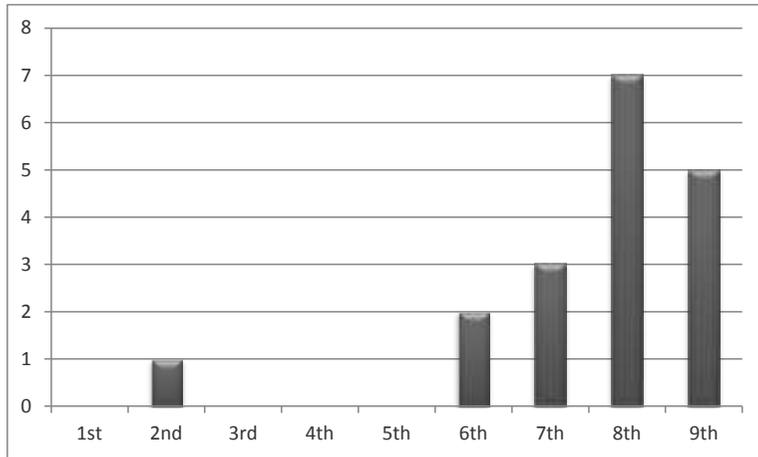
Topic C- HDR Brachytherapy in addition to External Beam Radiotherapy



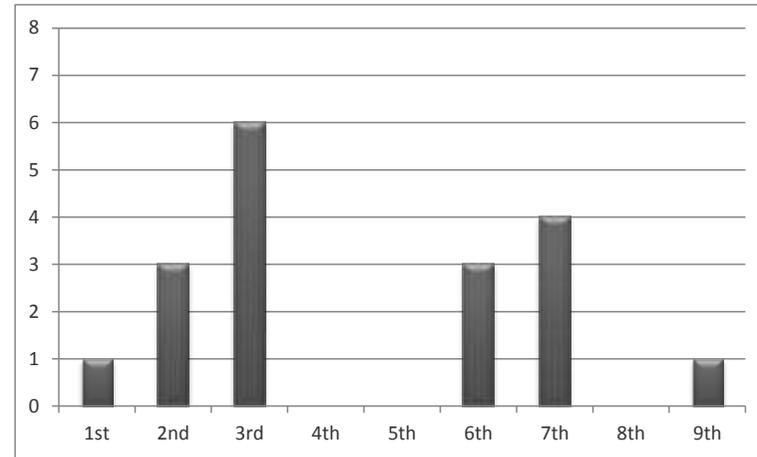
Topic D- LDR Brachytherapy in addition to External Beam Radiotherapy



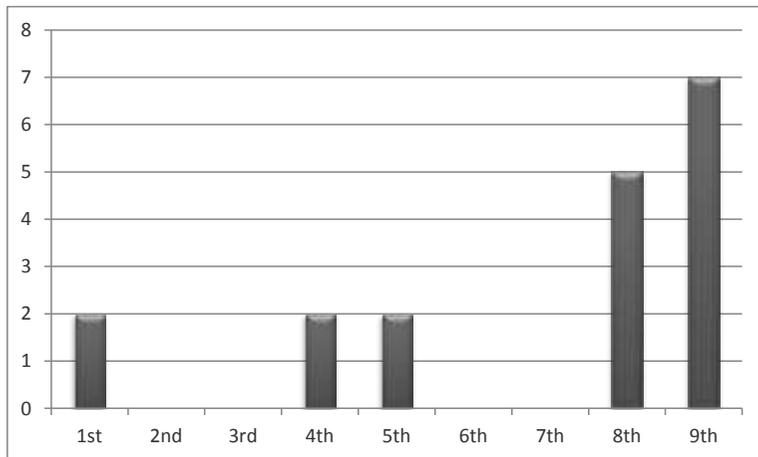
Topic E- Degarelix (a LHRH antagonist) for men with advanced hormone dependent prostate cancer



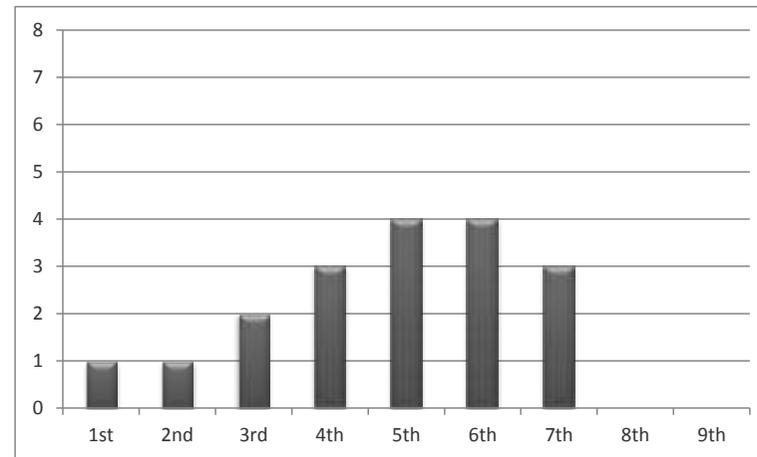
Topic F- Intermittent hormone therapy versus continuous hormone therapy



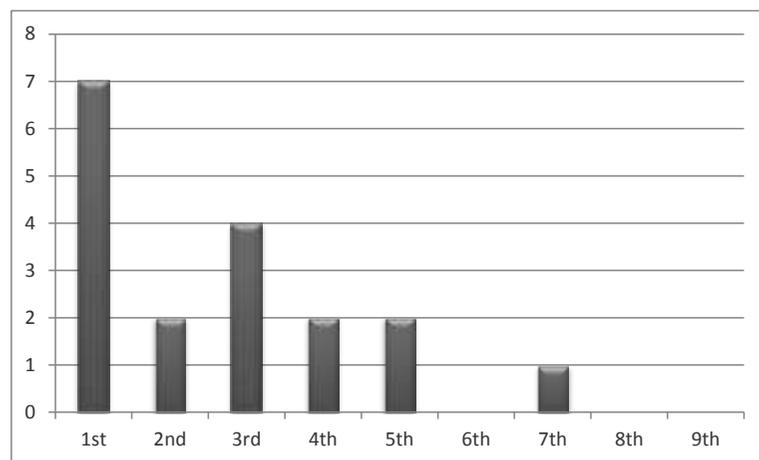
Topic G- Radium 223 chloride versus strontium-89



Topic H- IMRT and IGRT as an alternative to conventional radiotherapy

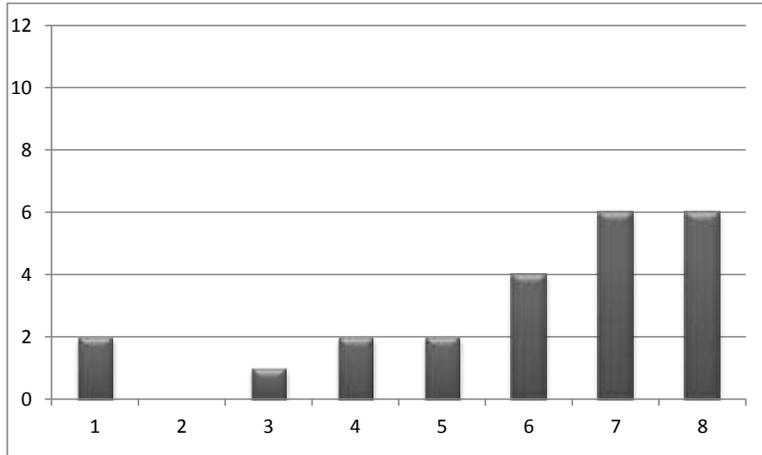


Topic I- Active surveillance in previously unscreened 'low risk' men

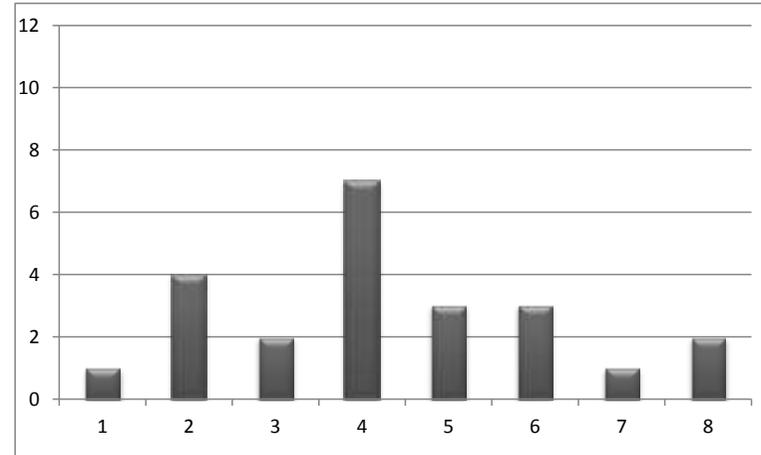


Atrial Fibrillation survey- Question 2 (Ranking of topics) n=23

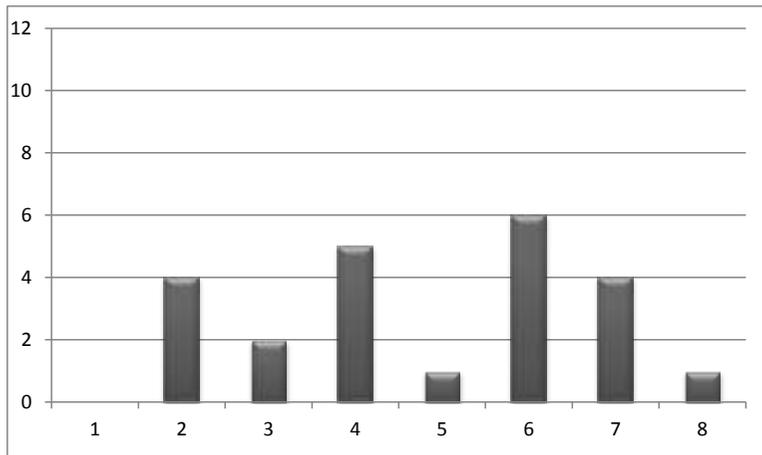
Topic A - Prophylaxis for the prevention of post-operative AF



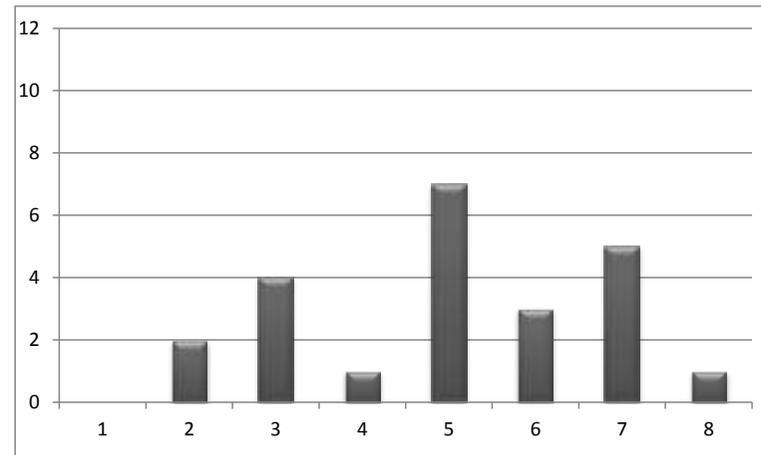
Topic B -Anti-arrhythmic drugs as pharmacological Cardioversion



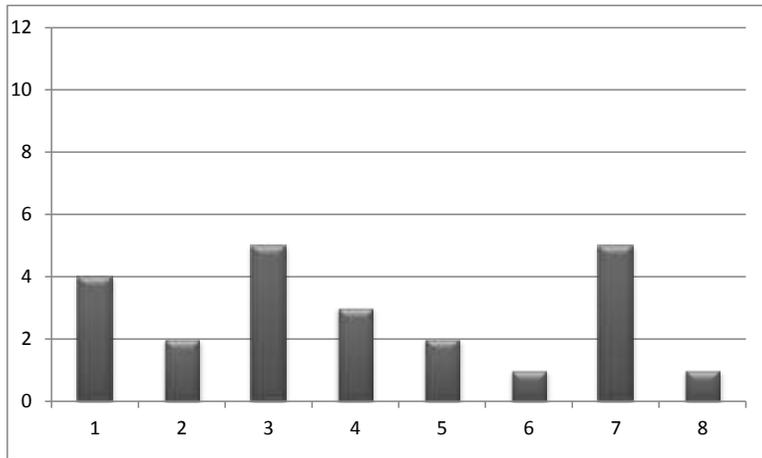
Topic C- Rhythm versus rate control strategies for persistent AF



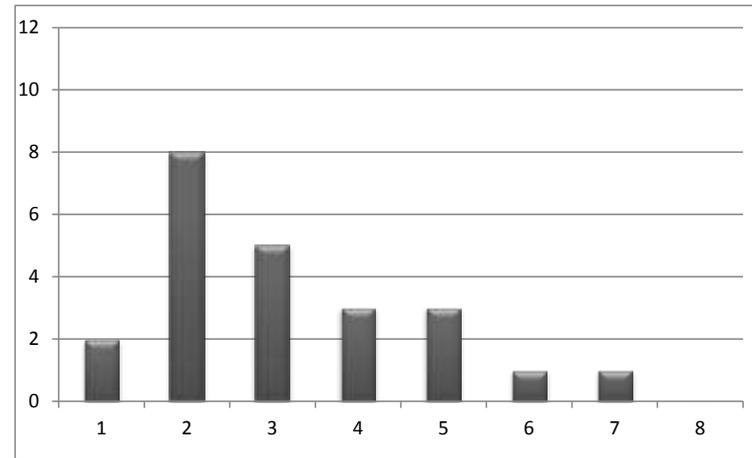
Topic D-Treatment for maintaining Sinus Rhythm after Cardioversion



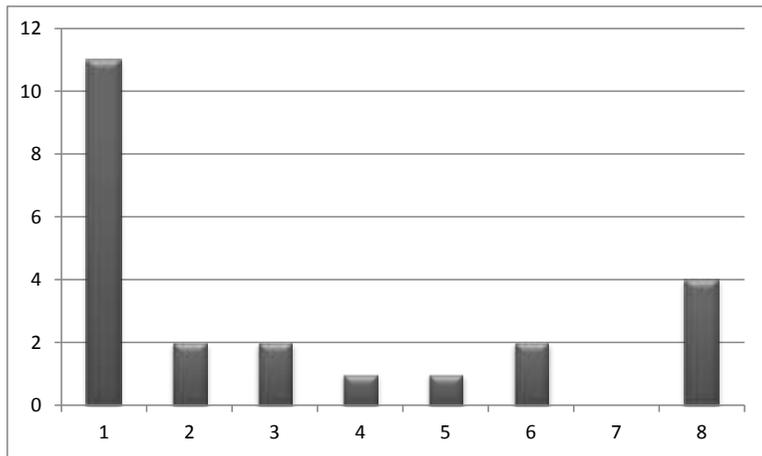
Topic E - Alternative Risk factor based scoring systems for stroke risk



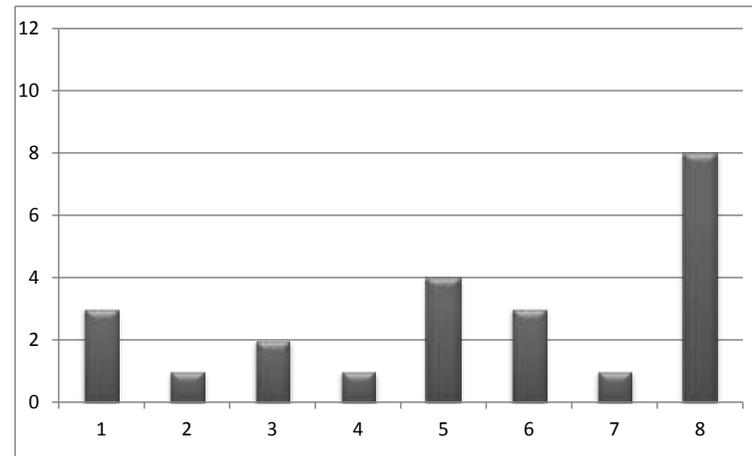
Topic F- Stratification tools to assess bleeding risk



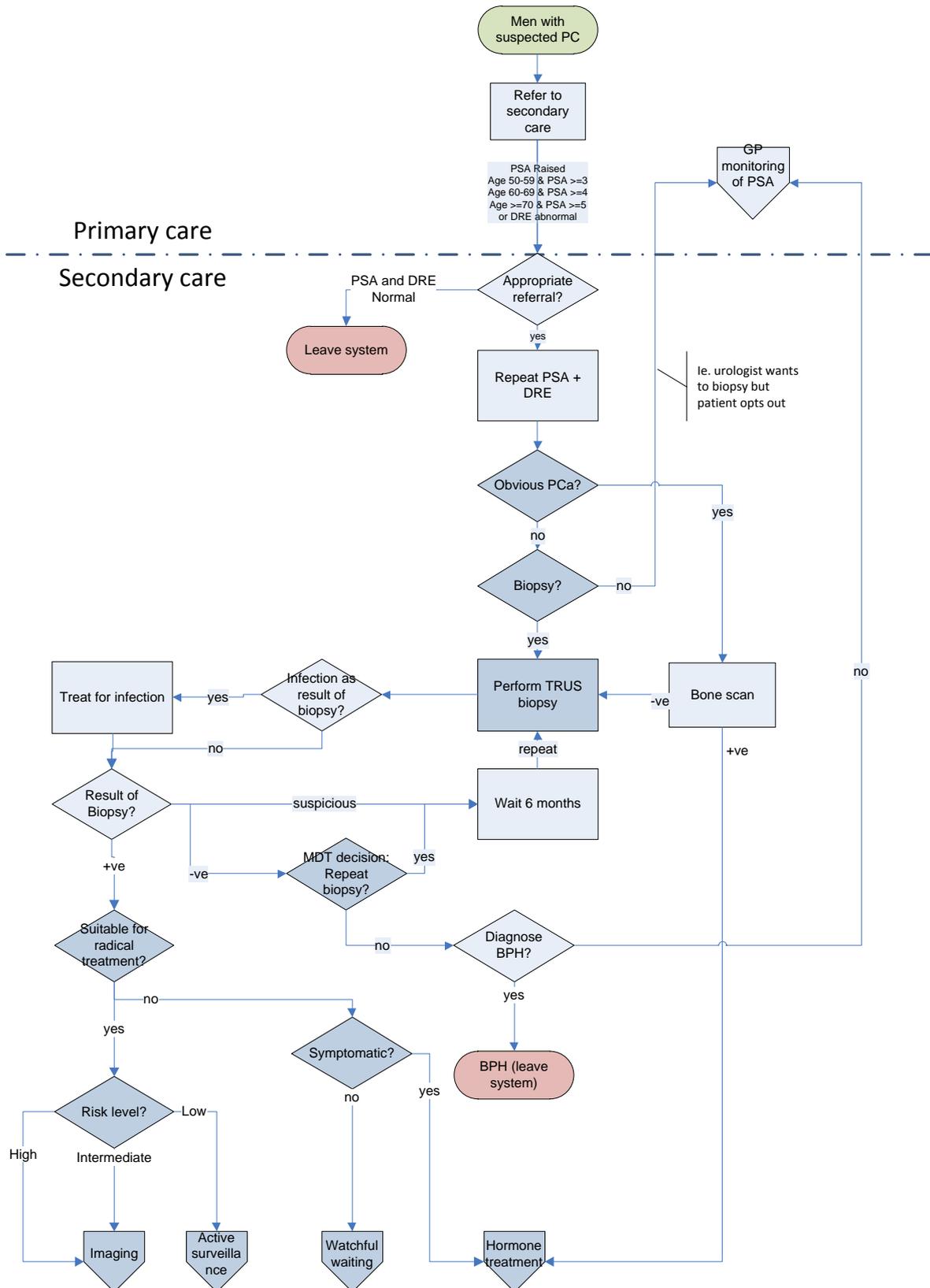
Topic G - Apixaban, rivaroxaban or Dabigatran etexilate

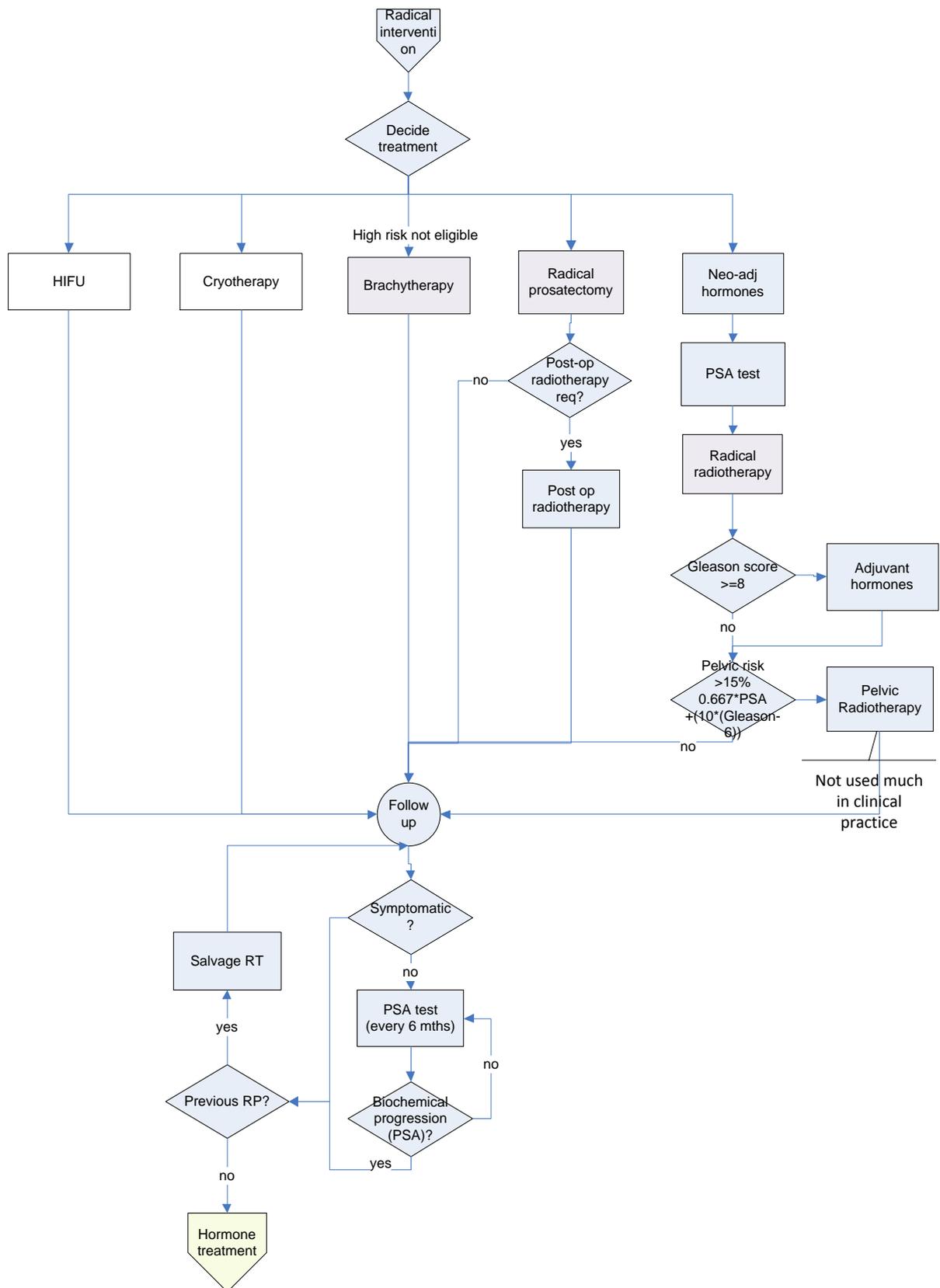


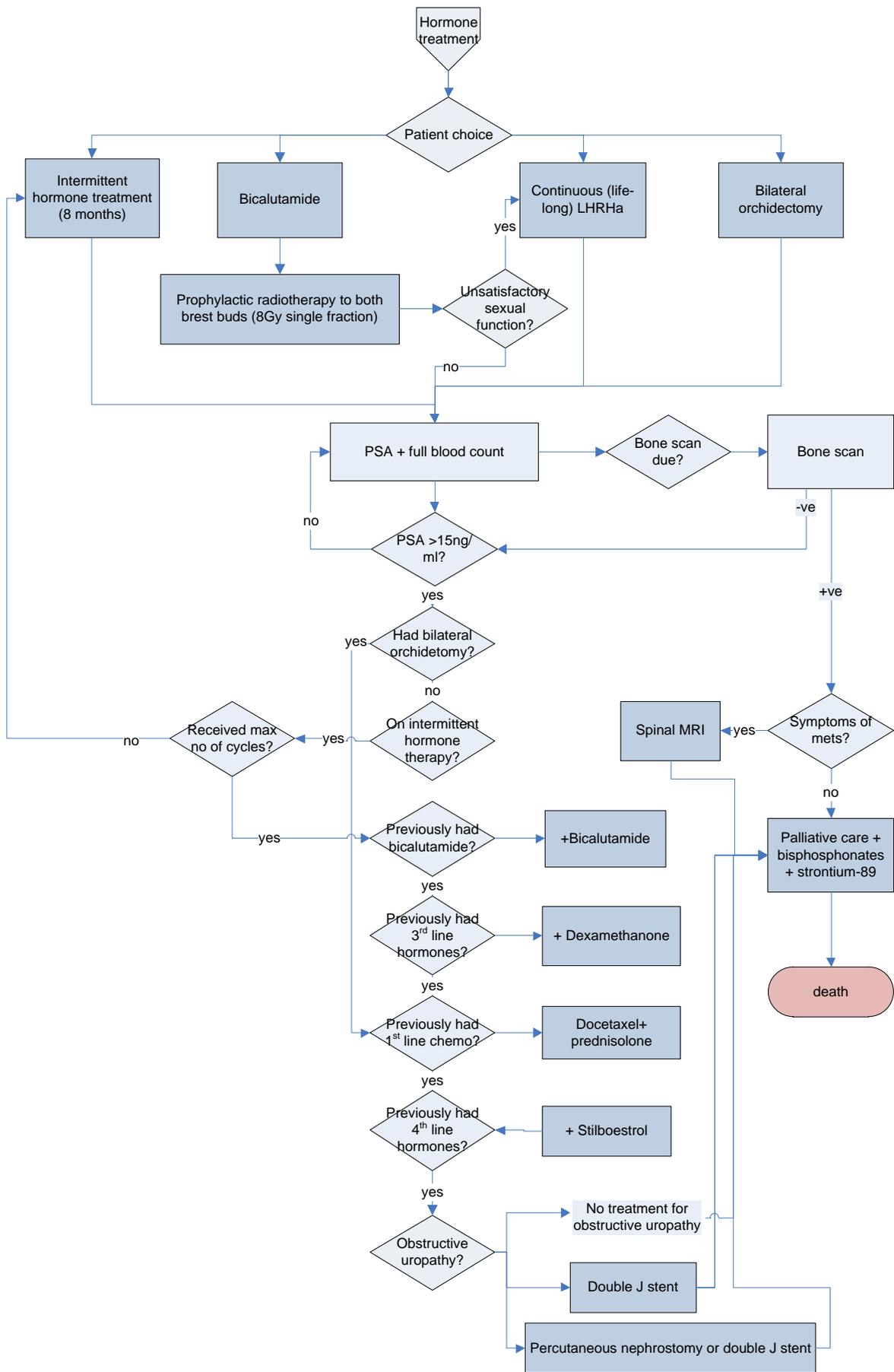
Topic H- Catheter ablation



Appendix 4. Prostate cancer service pathway







Appendix 5. Prostate cancer data for update topics

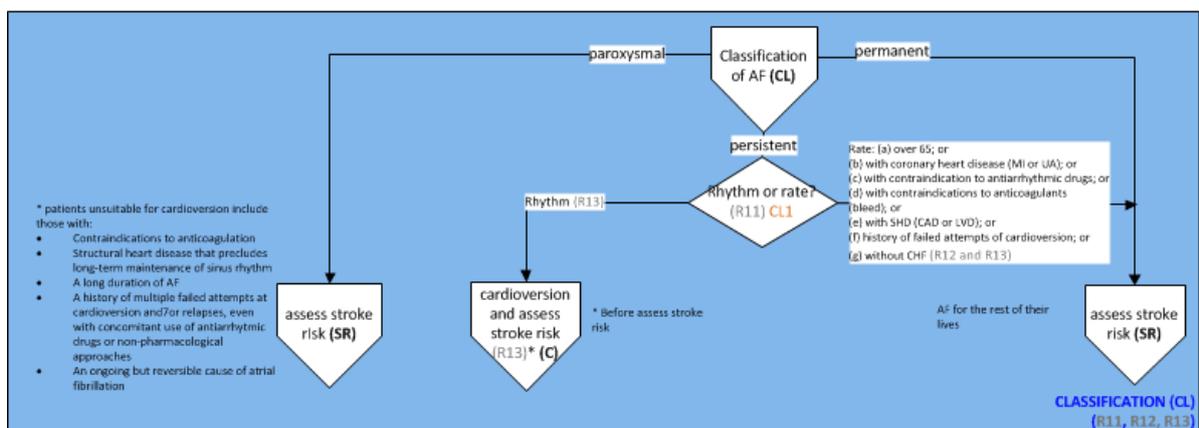
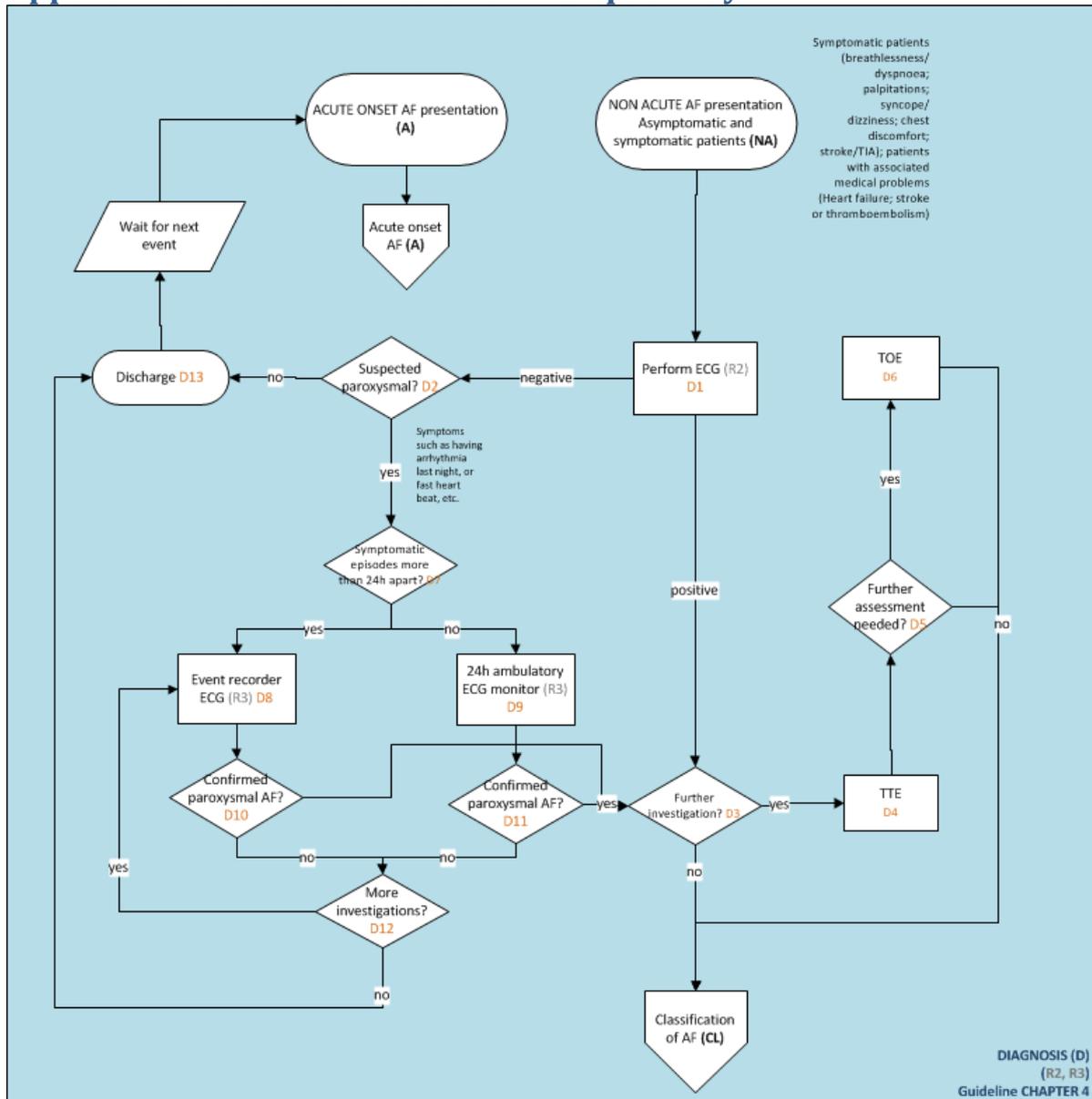
Topic	Treatment	Model parameter	First-order uncertainty	Second order uncertainty	Comments	Source
A	No new data					
B	Open radical retropubic prostatectomy (RRP)				Same as in base case model.	
	Transperineal prostatectomy (PRP)	Time to biochemical progression	Exponential ($\alpha=0.016$)	Normal ($\lambda= 0.016, se=0.002$)	Biochemical free survival curve same as for RRP (assumption using results of Martis <i>et al</i> 2007)	Bill-Axelson ⁷⁷ , Martis ¹⁰⁹
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha =70, \beta=30, mean = 0.70$)		Martis ¹⁰⁹
		Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha =26, \beta=74, mean = 0.26$)		
		Probability of bowel function AE	0	0	None reported assume no bowel related adverse events, as for RRP in base case model	Assumption
		Cost of PRP procedure	n/a	Cost of RRP varied as in base case. Cost of one bed day fixed (£267).	Cost of RRP (varied probabilistically as in base case model) minus the cost of one bed day (not varied in PSA).	NHS Reference costs ¹⁰⁴ . Assumption from Oxford Radcliffe Hospitals business case.
	Laparoscopic prostatectomy (LRP)	Time to biochemical progression	Exponential ($\alpha=0.016$)	Normal ($\lambda= 0.016, se=0.002$)	Use same biochemical free survival curve as above	Assumption, Bill-Axelson ⁷⁷
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha =46, \beta=18, mean = 0.72$)		Asimakpoulos ¹¹¹
Probability of urinary function AE		Uniform (0,1)	Beta ($\alpha =16, \beta=48, mean = 0.25$)			

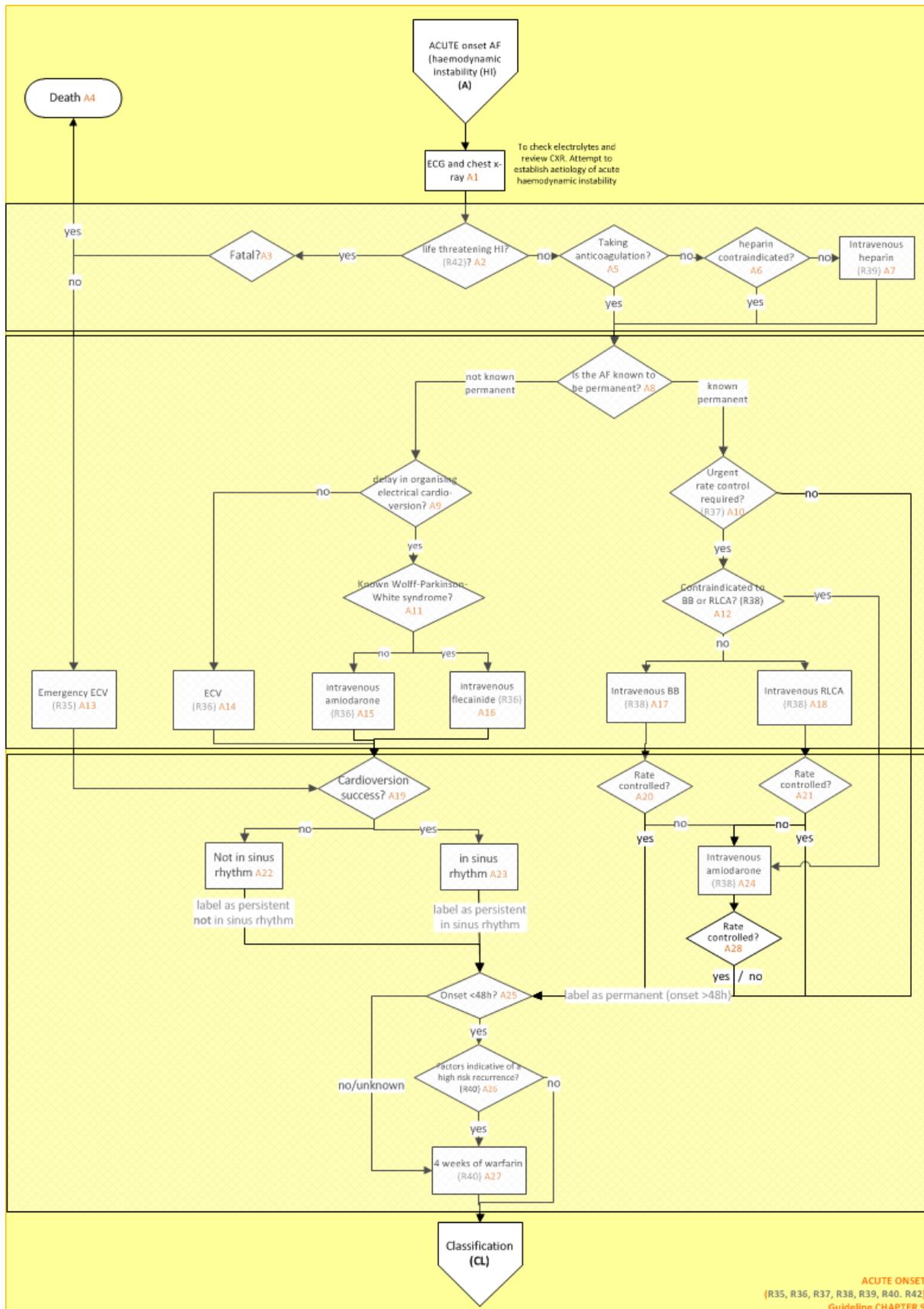
		Probability of bowel function AE	0	0	None reported assume no bowel related adverse events, as for RRP in base case model	Assumption
		Cost of LRP procedure	n/a	Normal (£5874, assumed se £185)	'Laparoscopic Bladder Neck Procedures – Male' HRG code LB22Z.	NHS Reference costs ¹⁰⁴
	Robot-assisted laproscopic prostatectomy (RALRP)	Time to biochemical progression	Exponential ($\alpha=0.016$)	Normal ($\lambda=0.016$, $se=0.002$)	Use same biochemical free survival curve as above	Assumption
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha=16$, $\beta=48$, mean = 0.25)		Asimakpoulos ¹¹¹
		Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha=8$, $\beta=56$, mean = 0.125)		
		Probability of bowel function AE	0	0	None reported assume no bowel related adverse events, as for RRP in base case model	Assumption
		Cost of RALRP (excluding capital costs)	n/a	Fixed £2,144	Cost of RRP procedure as per base case mode plus additional (fixed) cost of RALRP of £2144 (inflated to 2010-11 prices)	Oxford Radcliffe Hospitals business case
		Cost of robot (capital)	n/a	Normal (£3,000, se assumed 500)	Approximate value used, based on a dual console robot including service costs, based on a throughput of around 150 patients per year.	Ramsay ¹¹²
C/D	High dose rate (HDR) brachytherapy and external beam radiotherapy (EBRT)	Time to biochemical recurrence	Weibull ($\alpha=0.591$, $\beta=23.591$)	Multivariate normal ($\log \lambda=-1.855$, $\gamma=0.644$)		Sathya ¹¹³
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha=35$, $\beta=16$, mean = 0.69)		
		Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha=2$, $\beta=49$, mean = 0.04)		
		Probability of bowel function AE	Uniform (0,1)	Beta ($\alpha=4$, $\beta=47$, mean = 0.08)		
		Number of EBRT fractions	n/a	Fixed (point estimate = 37)	Fractionation schedule as per Sathya RCT.	
		Cost of HDR brachytherapy preparation	n/a	Normal (£1,039, se 0)	Preparation for interstitial brachytherapy SC55Z. Standard error estimated at zero since the reported lower and upper quartile values were the same.	NHS Reference costs ¹⁰⁴
		Cost of HDR	n/a	Normal (£3,956, se	Deliver a fraction of Interstitial Radiotherapy	

		brachytherapy delivery		780)	SC28Z.	
Low dose rate (LDR) brachytherapy and external beam radiotherapy (EBRT)	Time to biochemical recurrence	Weibull ($\alpha=1.067$, $\beta=23.209$)	Multivariate normal ($\log \lambda = -3.445$, $\gamma = 1.021$)			Sylvester ¹²⁰
	Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha = 35$, $\beta = 16$, mean = 0.69)	None of the three adverse events modelled were reported in Sylvester 2007. Instead assumed frequency of adverse events the same as for HDR-brachytherapy + EBRT (conservative assumption).		Assumption; Sathya ¹¹³
	Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha = 2$, $\beta = 49$, mean = 0.04)			
	Probability of bowel function AE	Uniform (0,1)	Beta ($\alpha = 4$, $\beta = 47$, mean = 0.08)			
	Number of EBRT fractions	n/a	Fixed (point estimate = 15)	Fractionation schedule as per Sylvester study.		Sylvester ¹²⁰
	Cost of LDR brachytherapy planning	n/a	Normal (£412, se £97)	Same as base case, assumed applicable for low dose rate brachytherapy. Preparation for interstitial brachytherapy, HRG code SC55Z.	Same as base case, assumed applicable for low dose rate brachytherapy Deliver a fraction of Interstitial Radiotherapy HRG code, SC28Z.	NHS Reference costs ¹⁰⁴
	Cost of LDR brachytherapy delivery	n/a	Normal (£383, se £196)			
E	No modelling					
F	No new data					
G	Radium-223 chloride	Additional overall survival	Weibull ($\alpha=2.173$, $\beta=10.582$) mean =9.371	Multivariate normal ($\log \lambda = -0.873$, $\gamma=1.696$)	<i>Note this topic was not evaluated since Radium-223 had no list price at the time of analysis.</i>	ALSYMPCA trial, personal communication. 28/06/12.
H	Intensity modulated radiotherapy (IMRT)	Time to local progression	weibull ($\alpha=1.354431605$, $\beta=21.78254729$)	Multivariate normal ($\log \lambda = -4.17$, $\gamma = 1.35$)	Scenario assumes no survival difference between IMRT and 3D-RT	Assumption; Hummel ¹⁰² ; Widmark ⁹⁴
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha = 250$, $\beta = 85$, mean = 0.75)	No difference between IMRT and 3D-RT. Assume probability of sexual function as for conformal radiotherapy.	
		Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha = 64$, $\beta = 289$, mean = 0.18)	No difference between IMRT and 3D-RT. Assume probability of sexual function as for conformal radiotherapy.	
		Probability of bowel function AE	Uniform (0,1)	Beta ($\alpha = 35$, $\beta = 110$, mean = 0.24)		Vora ¹²⁴

		Excess cost of IMRT (compared to 3D-RT)	n/a	Fixed (£1160)	Inflated to 2010-11 prices.	Hummel ¹⁰²
3D-conformal radiotherapy (3D-RT)		Time to local progression	weibull ($\alpha=1.354431605$, $\beta=21.78254729$)	Multivariate normal ($\log \lambda = -4.17$, $\gamma = 1.35$)	Scenario assumes no survival difference between IMRT and 3D-RT.	Assumption; Hummel ¹⁰² ; Widmark ⁹⁴
		Probability of sexual function AE	Uniform (0,1)	Beta ($\alpha = 250$, $\beta = 85$, mean = 0.75)	No difference between IMRT and 3D-RT. Assume probability of sexual function as for conformal radiotherapy.	Assumption; Widmark ⁹⁴
		Probability of urinary function AE	Uniform (0,1)	Beta ($\alpha = 64$, $\beta = 289$, mean = 0.18)	No difference between IMRT and 3D-RT. Assume probability of sexual function as for conformal radiotherapy.	
		Probability of bowel function AE	Uniform (0,1)	Beta ($\alpha = 37$, $\beta = 312$, mean = 0.1)	Assume probability of sexual function as for conformal radiotherapy.	
		Cost of 3D-RT planning	n/a	Normal (£581, se £81)	Same as in base case model. Preparation for complex conformal radiotherapy SC51Z.	NHS Reference costs ¹⁰⁴
		Cost of 3D-RT delivery (per fraction)	n/a	Normal (£111, se £6)	Same as in base case model. 37 fractions assumed. Deliver a fraction of complex treatment on a megavoltage machine SC23Z.	
	I	Not evaluated				

Appendix 6. Atrial Fibrillation clinical pathway





ACUTE ONSET
(R35, R36, R37, R38, R39, R40, R42)
Guideline CHAPTER 9

