

# **A Markov model for long-term cost-effectiveness modelling of screening for abdominal aortic aneurysms**

Version 2

July 2006

Lois Kim, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge  
CB2 2SR, UK (email: [lois.kim@mrc-bsu.cam.ac.uk](mailto:lois.kim@mrc-bsu.cam.ac.uk))

in collaboration with

Simon Thompson, MRC Biostatistics Unit, Cambridge, UK

Martin Buxton, Health Economics Research Group, Brunel University, UK

Andrew Briggs, Health Economics Research Centre, Oxford University, UK

Helen Campbell, Health Economics Research Centre, Oxford University, UK

## **Summary**

This report details the development of a Markov model for evaluating the long-term cost-effectiveness of screening for abdominal aortic aneurysms. It covers development of the model structure, model assumptions, parameter estimation, and internal validation. The model is based on a large randomised trial of screening, the Multicentre Aneurysm Screening Study (MASS), and its available data on mortality and costs up to an average of four years of follow-up. The principal purpose of this work is to develop a well-justified model suitable for long-term extrapolation.

# CONTENTS

- 1 Introduction
- 2 Development of the Markov model
  - 2.1 Model components
    - 2.1.1 AAA growth
    - 2.1.2 Elective surgery
    - 2.1.3 Emergency surgery
    - 2.1.4 Non-AAA mortality
    - 2.1.5 Incidental screening
    - 2.1.6 Loss to recall follow-up
  - 2.2 A proposed model structure
    - 2.2.1 Transition states
    - 2.2.2 Decision tree pathway
    - 2.2.3 Model population and transition parameters
    - 2.2.4 Model cost parameters
  - 2.3 Assumptions and simplifications: specific components
    - 2.3.1 AAA size and growth
    - 2.3.2 Consultation for considering elective surgery
    - 2.3.3 Elective and emergency surgery
    - 2.3.4 Incidental detection
    - 2.3.5 Non-AAA mortality
    - 2.3.6 Costs
  - 2.4 Assumptions and simplifications: general
    - 2.4.1 Markov modelling assumptions
    - 2.4.2 Censoring
    - 2.4.3 Timing of outcomes and costs
    - 2.4.4 Discounting

- 2.4.5 Sensitivity and specificity of the screening test
- 2.4.6 Parameter dependence on covariates
- 3 Estimation of parameters
  - 3.1 Estimation of individual model components from MASS
    - 3.1.1 Calculation of person-years used in parameter estimations
    - 3.1.2 Parameters defining distribution across starting states
    - 3.1.3 Rupture rates
    - 3.1.4 Dropout rates
    - 3.1.5 Aneurysm growth parameters
    - 3.1.6 Estimation of outcomes following referral
    - 3.1.7 Emergency surgery parameters
    - 3.1.8 Post-operative death rates
    - 3.1.9 Non-AAA death rates
    - 3.1.10 Cost parameters
  - 3.2 Estimation of components from non-MASS sources
    - 3.2.1 Growth in “non aneurysm” state
    - 3.2.2 Rate of incidental detection
- 4 Validation against four-year MASS results
  - 4.1 Total events, benefits and costs
  - 4.2 Events over time
    - 4.2.1 Elective operations
    - 4.2.2 Ruptures
    - 4.2.3 Emergency operations
    - 4.2.4 AAA deaths
    - 4.2.5 Non-AAA deaths
    - 4.2.6 Numbers in screen-detected AAA states

4.3 Conclusions and discussion

4.3.1 Model validation

4.3.2 Model development

5 References

## 1 Introduction

The Multi-centre Aneurysm Screening Study (MASS) has provided evidence on benefits and costs for abdominal aortic aneurysm (AAA) screening in men in the UK. The trial showed borderline cost-effectiveness of £28,400 per life-year saved (£36,000 per QALY) after four-years follow-up. These results and the hazard ratio showing significant benefit in terms of AAA-related mortality (0.52, 95% CI 0.42 - 0.78) have indicated that a national screening programme should be considered. In order to assist the decision-making process, the details for implementing such a programme must be optimised, and evidence for long-term cost-effectiveness provided.

A long-term model for AAA screening needs to incorporate aspects of aneurysm biology alongside aspects of the intervention, bringing together evidence on AAA growth, rupture and the associated mortality with ultrasound monitoring of the aneurysm, consultations for considering surgical interventions, and the surgical interventions themselves. The aim is to model the long-term costs and outcomes of a systematic screening programme versus a no systematic screening strategy. This would be evaluated over the lifetime of the individuals following entry into the screening programme in order to provide lifetime cost-effectiveness estimates. The model would aim to evaluate this based on observed aortic diameter measurements as opposed to the underlying true aortic diameters. Although it is known that there are inaccuracies in ultrasound measurements of the aorta, a model based on observed values is necessary since management of the AAA is determined by these. Some other models of the long-term cost-effectiveness have been published, but do not provide realistic estimates for the screening programme that is proposed; some make over-simplifications regarding the model structure, provide estimates for targeted screening only, do not use sufficient patient-level data to estimate costs, or do not carry out a probabilistic analysis allowing for joint parameter uncertainty.

In addition to cost-effectiveness estimates, a long-term screening model would also enable different fundamentals of the screening programme to be explored. A number of aspects of the implementation require further evidence, such as the age invited to attend screening, the screening of women, the use of alternatives to ultrasound in scanning (and

the associated sensitivity and specificity of such tests), possible re-screening of those with a non-aneurysmal aorta at initial scan, the definition of a non-aneurysmal aorta, length of intervals between recall scans, and the criteria for considering elective surgery. These details could be investigated by changes to the parameters or underlying structure of the model, and subsequent comparisons of cost-effectiveness of different screening strategies. The basic structure of the model must also incorporate the uncertainty about parameter estimates, which will initially be estimated from the MASS data, and then from systematic literature review. Uncertainty will be incorporated in the model by means of probabilistic analysis.

The development of the Markov model is fully described in Section 2, including a discussion of the model components, structure, parameters and assumptions. Section 3 details the estimation of the model parameters from the MASS data, and Section 4 uses these estimates in the model to check that they replicate the four-year results observed in the MASS trial. Model fit and sensitivity of outcomes to input parameters are discussed.

## **2 Development of the Markov model**

### **2.1 Model components**

A model of AAA screening must comprise a number of key components relating to the aneurysm development, monitoring and intervention. In order to develop a realistic model of the potential screening programme, the nature of each of these components must be understood in detail.

#### **2.1.1 AAA growth**

The size of an abdominal aortic aneurysm is an important indicator of the probability of rupture. It is therefore the focus of a screening programme to monitor aneurysm size over time, in order to assess the underlying risk of rupture and to intervene before this risk becomes too high. Coupled with this repeat screening strategy is the opportunity to predict when an aneurysm may reach a certain size, based on previous measurements. Aneurysms grow at different speeds, and it has been suggested that this growth may occur in spurts, and may depend on size as well as other patient characteristics. Aneurysm growth itself may also be an indicator of risk of rupture. It is important, therefore, to accurately reflect the growth of aneurysms in the Markov model, as well as the potential for aortas described as non-aneurysmal (<3.0cm) to grow and reach a size considered to be aneurysmal.

The measurement error associated with ultrasound scans of the aorta has been well documented, and creates a number of issues in any model of aneurysm growth. The error for a scan may be as high as 0.6cm, and hence at subsequent scans, the aortic measurement in the same patient may appear to have decreased. However, it is not believed that aortas truly shrink in size over time, but that they are only able to remain static, or to increase in size. This is dealt with in part by model simplifications employing categories of aneurysm size rather than modelling size as a continuous measurement. These categories are in practice related to the screening programme recall strategy, which in turn is based on evidence regarding aneurysm growth. The recall intervals

implemented in MASS correspond to three categories of aneurysm size; individuals with small aneurysms (described as an aortic diameter 3.0cm-4.4cm) received annual repeat scans, medium aneurysms (4.5cm-5.4cm) 3-monthly repeat scans, and large aneurysms ( $\geq 5.5$ cm) were considered for elective surgical intervention. However, the use of such categorisation in a progression-only model results in a degree of misclassification due to the measurement error. This issue has been investigated using a subset of data from a pilot trial for MASS based in Chichester, and requires the use of a hidden Markov model and estimation of state sojourn times. Incorporation of this work into the model would enable improved estimation of outcomes and investigation of optimising the screening strategy, but would require careful consideration of costs, as events in the screening programme occur based on observed measurements rather than the true underlying ones.

### **2.1.2 Elective surgery**

There is some uncertainty regarding the criteria that should be used to flag an individual for an appointment with a vascular consultant to consider elective surgery. However, current guidelines suggest using symptoms attributable to the aneurysm and a specified aortic diameter, although it is unclear where this cut-off should be. Different AAA screening programmes have implemented various size criteria, ranging from 5.0cm to 6.0cm. The consequences of operating at even smaller diameters has been investigated in two randomised trials, and found not to be beneficial in terms of AAA-related mortality. The MASS trial used a 5.5cm indicator, and additionally a “rapid expansion” criterion, indicating a consultation when the aneurysm grew  $\geq 1.0$ cm over a one-year period. The role of this criterion as a useful indicator of rupture remains uncertain, but recent recommendations suggest it is of little additional benefit in a screening programme, since the risk of mortality from rupture in a given subgroup should be greater than the risk of mortality from the intervention in order for the intervention to be justified.

Once a criterion has been met, the patient may receive one or more consultations with a vascular surgeon. A small number of referred patients may refuse the consultation, while others may require further tests and extra appointments, so the number of consultations per individual varies. At the consultation the aorta is rescanned and, due to the

measurement error associated with ultrasound scans, the new measurement may no longer meet the size criterion. These patients may then be returned to the recall screening cycle. Patients confirmed as meeting the size criterion at the consultation may either be declared unfit for elective surgery, refuse the operation, or decide to go ahead with it. An unsuccessful outcome following elective surgery is defined as all-cause 30-day mortality (post-operative mortality) after the AAA repair. Fitness for and outcome following elective surgery are dependent on a number of factors, including age.

### **2.1.3 Emergency surgery and AAA rupture**

In the absence of elective surgical intervention, an aneurysm may continue to grow and eventually rupture. The risk of rupture increases with increasing aneurysm size, and may also be related to aneurysm growth rates. A number of predictors of aneurysm rupture have also been documented, including age. In the event of the rupture of an aneurysm, emergency surgery carrying out AAA repair is necessary to prevent death. The proportion of AAA rupture patients reaching hospital for this vital surgery, however, is low. Furthermore, survival following such surgery is also low, hence the benefit of identifying and operating electively on high-risk aneurysm patients via a screening programme. An unsuccessful emergency AAA operation is similarly defined by post-operative all-cause mortality within 30 days of surgery.

### **2.1.4 Non-AAA mortality**

The MASS trial assessed the benefit of aneurysm screening in terms of AAA-related mortality. This was defined as all deaths recorded on the death certificate as abdominal aortic aneurysm deaths (ICD9 codes 441.3, 441.4), aortic aneurysm deaths at an unspecified site (ICD9 codes 441.5, 441.6), and 30-day all-cause mortality following any surgery for AAA repair. In the Markov model, the probability of AAA-related mortality from each of the non-absorbing states can be estimated from published data and data from MASS. Deaths from other causes may be modelled using age-standardised national mortality statistics, although patients with an aneurysm are known to have other cardiovascular co-morbidities and a relatively high mortality. Within this group, patients

assessed for elective surgery and declared unfit have an even higher risk of all-cause mortality. There is additionally uncertainty regarding the similarity of death rates to national standardised death rates following successful aneurysm repair. Evidence suggests that individuals not attending for a scan following invitation to screening also have a higher all-cause mortality than the general population. Consequently, these subgroups may require separate estimation of all-cause mortality.

The population recruited into the MASS trial are known to be less socially deprived than the median for England and Wales, which has implications for the application of national death rates to the model when checking the four-year results from the trial. Therefore the national death rates are only applied for the lifetime model, and non-AAA deaths rates taken directly from the trial population are used instead in the four-year model.

### **2.1.5 Incidental screening**

Since ultrasound is a quick, non-invasive procedure used for a number of other purposes, it is to be expected that some incidental screening will take place in conjunction with other medical investigations outside of systematic screening. This will occur in the population in addition to any systematic screening programme being implemented, as incidental screening may take place in non-attenders. In the control group in the MASS trial (i.e. those not invited to attend the screening programme), elective operations resulting from an incidental scan were recorded, but data regarding incidental scanning itself was not collected. The available published evidence is also sparse, particularly with respect to the amount of incidental screening that may be occurring in the community as a whole, rather than in specific subgroups. It is also unclear whether the probability of detection varies according to the aneurysm size (which may in turn be related to other co-morbidities that instigate the incidental scanning). However, it is necessary in the model to estimate the proportion of initially undetected aneurysms that are detected by incidental screening since this has implications for both the costs and outcomes being investigated. The overall cost-effectiveness of systematic AAA screening estimated by the model will obviously become diluted in the presence of increasing incidental screening.

### **2.1.6 Loss to recall follow-up**

During the course of the MASS trial, a significant number of individuals with identified aneurysms failed to attend for a scheduled recall scan. This loss to follow-up continued throughout the four-year follow-up period in MASS, principally from the small and medium size groups. Those with an identified large aneurysm more commonly refused an elective operation, which can be included in the model as contraindication for elective surgery (see 2.1.2); consequently, loss from the recall cycle is not modelled for this group. It is also important to model those lost from the recall scan cycle in the small and medium size groups, both in terms of costs and numbers at risk for events such as rupture.

## **2.2 A proposed model structure**

The overall model must estimate costs and outcomes for both systematic screening and no systematic screening. The basic model structure is therefore two-arm, replicating the same structure, for each of these scenarios. Although the two arms described here have identical structures, some of the transition states must be interpreted differently. Furthermore, there are numerous different potential structures and levels of complexity that could be used to model AAA screening; the model proposed here is based on the MASS screening strategy, and makes a number of assumptions (detailed in Section 2.3). The transition states of this model and decision nodes defining the pathways between them are described in detail below.

### **2.2.1 Transition states**

The model is initially populated by distributing the simulated individuals across a number of starting states according to parameters defining the probability of being in each of these states. The possible starting states in this model are defined in the screening arm by attendance and aneurysm size at the initial scan; this implies that screening occurs prior to the start of cycle 1. It is assumed that no incidental scanning has occurred prior to the start of the model in either arm; this assumption will be tested in later sensitivity analyses. Each starting state is described below, with distinctions between screening and no-screening arms made where necessary.

### *1. No AAA*

An aortic diameter is chosen to define a non-aneurysmal (“normal”) aorta. The model uses the <3.0cm definition employed in MASS, although other screening programmes have used <2.6cm. In the screening arm of the model, this starting state includes three subgroups; those attending the initial scan and known to have a normal aorta, those attending the initial scan but with unsuccessful aorta visualisation of a normal aorta (1% of all scanned cannot be successfully visualised), and those not attending the initial scan but having a normal aorta. Some assumption must be made concerning the expected proportions of non-attenders and non-visualised attenders with a normal aorta, which may or may not be the same as the proportion of attenders with a normal aorta. In the no-screening arm, the no AAA starting state simply includes all those in the population with an aortic diameter <3.0cm.

### *2. Small AAA, detected*

A small AAA is defined as being an aortic diameter 3.0-4.4cm, as used in MASS. In MASS, small AAAs detected at the initial scan underwent recall screening on a yearly basis. In the screening arm, this group comprises attenders with a small AAA detected at screening. In the no-screening arm, since it is assumed that no individuals have received incidental screening before the start of the model, there are no individuals in this starting state.

### *3. Medium AAA, detected*

A medium AAA is defined as an aortic diameter 4.5-5.4cm. In MASS, detected aneurysms of this size underwent recall screening every three months. Similar notes apply as to the small-detected AAA state.

#### *4. Large AAA, detected*

A large AAA is defined as an aortic diameter  $\geq 5.5$ cm. In MASS, this was chosen as the size cut-off for considering elective surgery. All patients with a detected AAA reaching this size were referred for a consultation with a vascular surgeon. Similar notes apply as to the small-detected AAA state.

#### *5. Small AAA, undetected*

In the screening arm, this group comprises non-attenders and non-visualised attenders with a small AAA, aortic diameter 3.0-4.4cm. Since the proportions of non-attenders and non-visualised attenders with an aneurysm of this size is unknown, some assumption must be made with respect to the expected proportions, and whether this can be estimated from the proportion of attenders with a small detected AAA. In the no-screening arm, since it is assumed that no individuals have received incidental screening before the start of the model, all individuals with a small AAA at the start of the model begin in this undetected state.

#### *6. Medium AAA, undetected*

Similarly, this group comprises non-attenders and non-visualised attenders with a medium AAA (aortic diameter 4.5-5.4cm) in the screening arm, and all those starting with a medium AAA in the no-screening arm.

#### *7. Large AAA, undetected*

For the screening arm, this group comprises non-attenders and non-visualised attenders with a large AAA (aortic diameter  $\geq 5.5$ cm), and for the no screening arm, all those starting with a large AAA.

Following allocation to a starting state, each individual passes through a number of decision nodes relating to the aneurysm development and screening strategy, before

finishing the cycle in one of a number of states. This may be one of the seven states listed above, or one of six additional (non-starting) states:

*1. Emergency surgery survivor*

Any individual undergoing successful emergency surgery of a ruptured AAA (i.e. those surviving at least 30 days post-surgery) will enter this state. The individual then remains in this state for all subsequent cycles until death.

*2. Elective surgery survivor*

Similarly, any individual undergoing successful elective surgery of a non-ruptured AAA enters this state, and remains in this state for all subsequent cycles until death.

*3. Elective surgery pending*

Where elective surgery is indicated at consultation, the individual enters this state at the end of the cycle in large detected AAA. At the end of one cycle in the ‘elective surgery pending’ state, elective surgery has taken place, unless non-AAA death or rupture occurs first. This enables a delay of one cycle length between consultation and elective surgery to be incorporated in the model.

*4. Elective surgery contraindicated or refused*

This group comprises individuals with a large, detected AAA who have had a consultation for considering elective surgery, but have been declared unfit for this surgery, or refused to proceed with it.

*5. Dead from AAA*

This includes individuals dying from ruptured AAA without receiving emergency surgery in addition to all-cause 30-day mortality following any (emergency or elective) AAA repair. This is an absorbing state.

*6. Dead from other causes*

An absorbing state for all other causes of death not covered by the “dead from AAA” definition.

After the initial distribution of individuals across starting states, the states for small, medium and large detected AAAs include individuals with an aneurysm detected via incidental screening (in addition to those detected via the systematic screening programme in the screening arm). In order to obtain lifetime outcomes and costs, the model continues to run cycles until all individuals are in an absorbing state; this occurs when all individuals are in a death state. The length of each cycle is selected to be a length of time over which an individual can be known to change states, but also takes into account computational considerations. In the MASS screening programme, this is most easily represented by the three-monthly recall scans for medium AAAs; hence a cycle length is defined as three months. In reality, some other transitions between states (such as rupture) may occur over a shorter period than three months. However, the use of any cycle length makes assumptions regarding the timing of events, and the use of probabilities of events over the three-month cycle enables reasonable estimation of outcomes and costs. The principal difficulty lies with modelling the decision at consultation for elective surgery, since this cannot be estimated as a probability over a time period.

### **2.2.2 Decision tree pathway**

From each state at the start of a cycle, a number of decision nodes are passed through before the individual reaches the finishing state for the given cycle. There are a large number of possible combinations of events that could occur within a three-month period, and some simplification is necessary to create a model that contains estimable parameters. This is achieved by using consecutive decision nodes that essentially make certain events mutually exclusive. For example, although in reality an aneurysm could increase in size from small to medium and rupture in the same three-month period, this is unlikely. Instead, the model uses two separate decision nodes for rupture, then growth if no rupture is observed. This modelling also reflects the recall screening strategy, since any growth prior to a rupture within a three-month period following a scan would not be

detected before the rupture occurred. The influence of the choice of structure on the life-years and costs outcomes of the model can be investigated later through examination of the results of some alternative structures

The decision pathway for the starting state “no AAA” is given in Fig. 2.1, showing all the possible pathways of a cycle for an individual starting in the no AAA state.

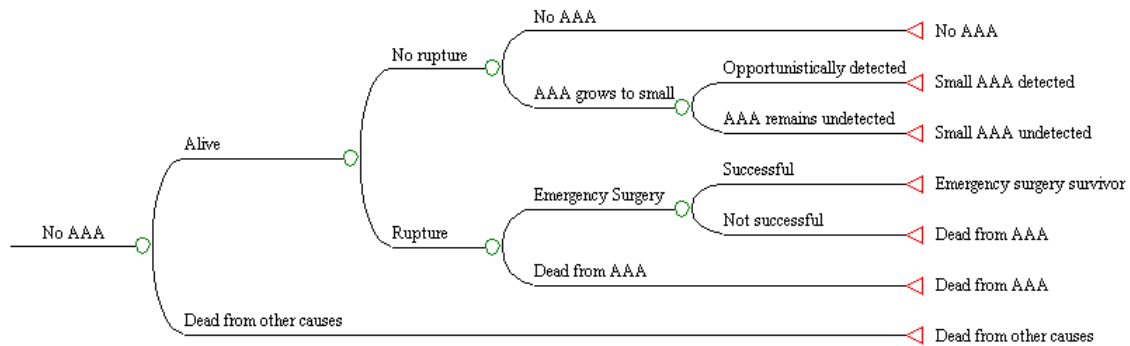


Fig. 2.1 Decision pathway for “no AAA” starting/ transition state

A probability is assigned to each outcome from a decision node, conditional on reaching that node. At the end of a cycle, an individual starting in the no AAA state finishes in one of six possible states, including remaining in the no AAA state. Individuals then enter the next starting state based on the Markov assumption that the past history (length of time remaining in a given state and pathway to getting there) has no influence on the parameters of the pathway. In this pathway for the no AAA state, there are only costs associated with emergency surgery, and with the initial scan for attenders in the screening arm when populating the starting states. Costs are not accrued for an incidental scan in this pathway, but only for subsequent recall scans associated with the aneurysm detected starting states. The model for small, medium and large undetected AAAs follows the same structure (although the latter cannot grow further, since there is no upper boundary to the definition of a large AAA;  $\geq 5.5\text{cm}$ ), additionally allowing for incidental detection from the starting size.

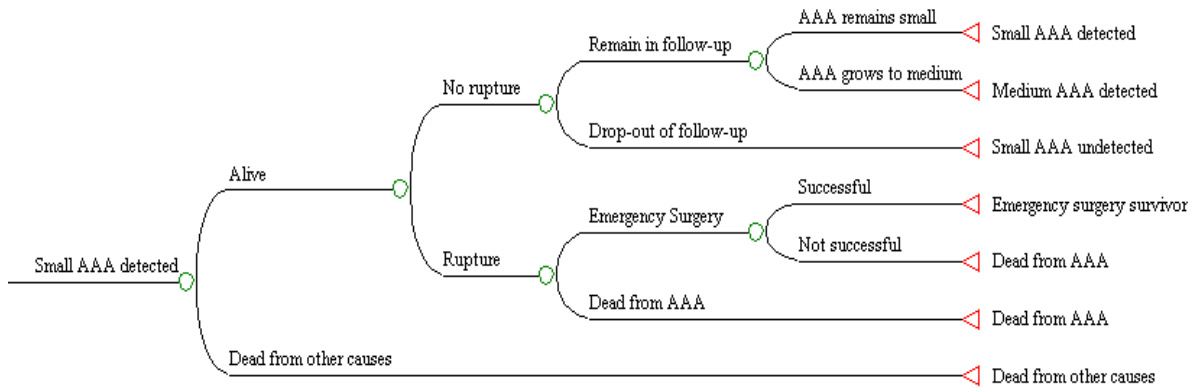


Fig. 2.2 Decision pathway for small AAA detected starting/ transition state

The pathway for the state “small AAA, detected” is similar (Fig. 2.2), excluding the nodes for incidental detection and adding nodes for loss to recall follow-up; the model for medium detected AAAs is similar.

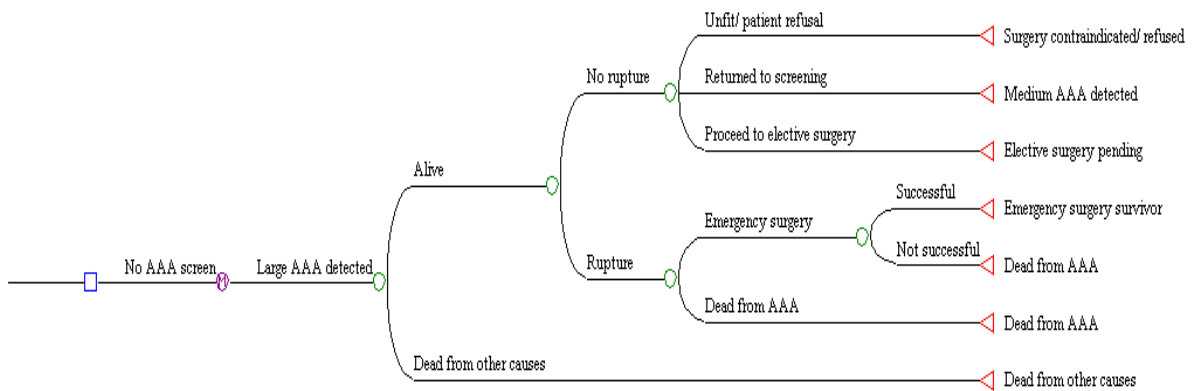


Fig. 2.3 Decision pathway for large AAA detected starting/ transition state

The pathway for the large detected AAA starting state (see Fig. 2.3) is different in that it includes the decision made at the consultation for considering elective surgery. This decision is represented as a 3-outcome node given no rupture occurs, which models the three principal outcomes of a consultation; unfit for surgery/ patient refusal of operation, return to recall screening owing to no confirmation of the size criterion, and decision to

proceed with elective surgery. In practice, the delay between the decision and an elective operation means that rupture can take place between these two events as well as between reaching large size and receiving a consultation. In MASS, the median delay between reaching the large size state (including those returned to this state) and the final consultation was 71 days (n=435), with five ruptures observed whilst the patient was awaiting a consultation or following delays in a decision; the median delay between first consultation and receiving elective surgery was 59 days (n=268), with nine ruptures observed after a decision for elective surgery was made. These delays are broadly accounted for by the ‘elective surgery pending’ state, which means the average time from reaching the large detected size state and surgery in the model is 90 days (two 3-month cycles with half-cycle correction, see Section 2.3.8).

Individuals entering the surgery contraindicated/ refused state cannot later undergo elective surgery in this model, but only emergency surgery following rupture. Although in reality a small number of these individuals may be reconsidered for and proceed with elective surgery, this seems preferable to having a separate, poorly estimated parameter for elective surgery in the surgery contraindicated/ refused state.

The full decision model showing transitions between all states for both screening and no screening is shown in Fig. 2.4. The AAA groups inside the grey boxes plus the “no AAA” state represent the seven starting states (Section 2.2.1). Ovals represent model states; rectangles are events that occur as part of a transition to another state. Numbers corresponding to parameters are shown for each transition (see Section 2.2.3 for full descriptions of both transition and starting state parameters). The transition parameter relating to non-AAA deaths is not indicated in the figure, since it is not represented by a single parameter estimate as all the other transition parameters are. Instead, it is age-dependent, and hence the estimate applied depends on the cycle number. This cycle-dependent estimate additionally applies to each of the states shown in the figure.

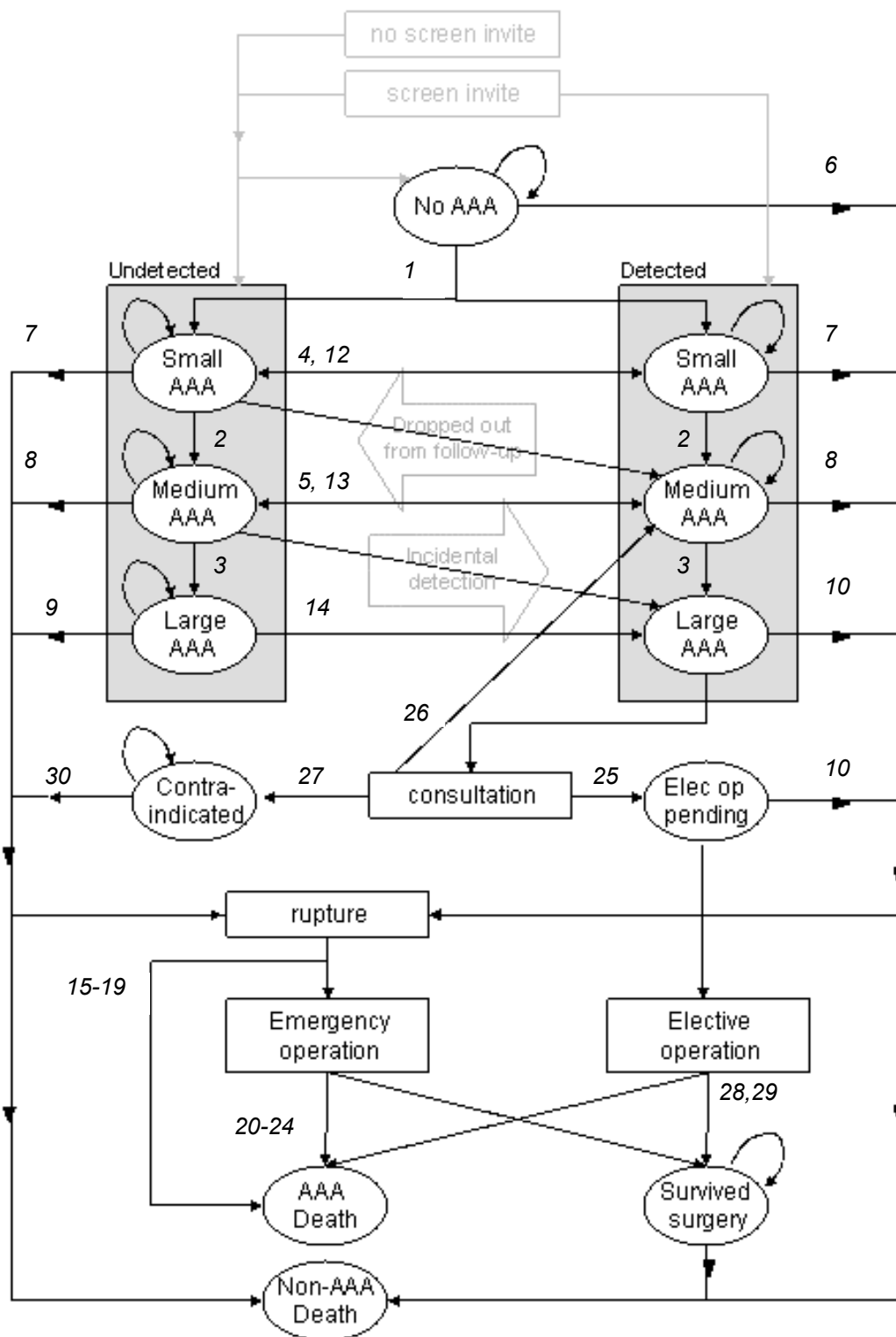


Fig. 2.4 Full Markov model, showing screening invitation and no screening arms

### 2.2.3 Model population and transition parameters

Each outcome from a decision node in the model is estimated by a parameter giving the probability of the event conditional on reaching that point in the pathway (see Table 2.2). Additionally, parameters describing the distribution of individuals across starting states are required (Table 2.1). The parameters will later be estimated using the MASS data and other available evidence, and are named and defined here for convenience.

Parameter name	Description <sup>+</sup>
<i>No screen</i>	
ns_grp1	No AAA
ns_grp2un	Small, undetected AAA
ns_grp3un	Medium, undetected AAA
ns_grp4un	Large, undetected AAA
ns_grp2det	Small AAA detected incidentally before start of model (assumed zero)
ns_grp3det	Medium AAA detected incidentally before start of model (assumed zero)
ns_grp4det	Large AAA detected incidentally before start of model (assumed zero)
<i>Screening invitation</i>	
reinvite*	Requiring re-invitation
attend**	Attending initial scan
nonvis**	Proportion of attenders with non-visualised scan
aneu_nv**	Proportion of non-visualised attenders with aneurysm (any size)
aneu_na**	Proportion of non-attenders with aneurysm (any size)
s_grp1	No AAA (attenders with no AAA, plus non-attenders and non-visualised attenders with no AAA)
s_grp2un	Non-attenders or non-visualised attenders with small AAA
s_grp3un	Non-attenders or non-visualised attenders with medium AAA
s_grp4un	Non-attenders or non-visualised attenders with large AAA
s_grp2det	Attenders with detected small AAA
s_grp3det	Attenders with detected medium AAA
s_grp4det	Attenders with detected large AAA

<sup>+</sup> Refers to proportion of whole control or invited population unless otherwise specified

\* Used in calculation of costs only

\*\* Used in calculation of other parameters only; not direct parameters in the model

Table 2.1 Parameters relating to model starting states

Parameters defining transition probabilities for given cycle length (same for no screen and screening invitation arms, unless otherwise stated):

<b>Transition number</b>	<b>Abbreviated parameter name</b>	<b>Description (3-month probability)</b>
<i>Growth</i>		
1	grow_grp1grp2	Aneurysm grows from <3.0cm to 3.0-4.4cm
2	grow_grp2grp3	Aneurysm grows from 3.0-4.4cm to 4.5-5.4cm
3	grow_grp3grp4	Aneurysm grows from 4.5-5.4cm to ≥5.5cm
<i>Loss to follow-up</i>		
4	lost_grp2	Lost to recall follow-up if small AAA
5	lost_grp3	Lost to recall follow-up if medium AAA
<i>Ruptures</i>		
6	rr_grp1	Rupture if no AAA (<3.0cm)
7	rr_grp2	Rupture if small AAA
8	rr_grp3	Rupture if medium AAA
9	rr_grp4un	Rupture if large, undetected AAA
10	rr_grp4det	Rupture if large, detected AAA
11	rr_contra	Rupture if contraindicated/ refused elective surgery
<i>Incidental detection</i>		
12	opp_grp2	Incidental detection of AAA if small (for both starting cycle as small, and growing to small during cycle)
13	opp_grp3	Incidental detection of AAA if medium (for both starting cycle as medium, and growing to medium during cycle)
14	opp_grp4	Incidental detection of AAA if large (for both starting cycle as large, and growing to large during cycle)
<i>Emergency surgery</i>		
15	emer_grp1	Emergency surgery if rupture with no AAA (<3.0cm)
16	emer_grp2	Emergency surgery if rupture with small AAA
17	emer_grp3	Emergency surgery if rupture with medium AAA
18	emer_grp4	Emergency surgery if rupture with large AAA
19	emer_contra	Emergency surgery if rupture after contraindicated/ refused elective surgery
20	deademer_grp1	Post-operative mortality if emergency surgery for rupture with no AAA (<3.0cm)
21	deademer_grp2	Post-operative mortality if emergency surgery for rupture with small AAA
22	deademer_grp3	Post-operative mortality if emergency surgery for rupture with medium AAA
23	deademer_grp4	Post-operative mortality if emergency surgery for rupture with large AAA
24	deademer_contra	Post-operative mortality following emergency surgery for rupture after contraindicated/ refused elective surgery
<i>Elective surgery &amp; decisions at consultation</i>		
25	elec	Consultation outcome is proceed with elective surgery
26	return	Consultation outcome is return to screening
27	contra	Consultation outcome is contraindication/ refusal of elective surgery
28	ns_deadelec	Post-operative mortality following elective surgery via incidental detection
29	S_deadelec	Post-operative mortality following elective surgery via screening programme detection
30	dead_contra	All cause mortality if contraindicated for/ refused elective surgery

Table 2.2 Parameters relating to model transitions

The rupture rate amongst individuals with a large AAA is estimated by two separate parameters for detected and undetected aneurysms. This reflects the difference resulting from intervention for the majority of large detected AAAs; other AAA sizes associated with no intervention can be assumed to have similar risk of rupture, unless it is considered that screening itself results in lifestyle changes that reduce the risk of rupture. The rupture rate in undetected large AAAs may over time become more similar to the rupture rate in detected large AAAs contraindicated for elective surgery, as a result of the lack of intervention. In the absence of direct evidence about undetected large AAAs, the rupture rate in this group may be best estimated from data from large AAAs contraindicated for elective surgery.

Post-operative mortality following elective surgery is also estimated by two separate parameters for operations following identification of the AAA via systematic or incidental screening. Although strictly this violates the Markov assumptions relating to past history, AAA states following incidental detection could be thought of as separate from those following systematic screening. All the transition parameters would be identical, with the exception of the post-operative mortality following elective surgery. The estimates may be expected to be different due to the likely presence of comorbidities in those with incidentally detected aneurysms, which often resulted in the initial scan taking place.

With respect to the estimation of starting state probabilities, three additional parameters (“attend”, “aneu\_nv” and “aneu\_na”) are used in the calculation of starting state parameters for the model. The “attend” parameter represents the proportion of those invited to screening who attend the initial scan. In MASS, this figure was 80%, and this is used to estimate this parameter in the baseline model. However, changing this parameter enables investigation of the sensitivity of results to different levels of uptake of screening, based on the assumption that the proportion and distribution of aneurysms remains the same in attenders. Following on from this, “aneu\_na” represents the proportion of non-attenders with an aneurysm of any size, and similarly “aneu\_nv” the proportion of non-visualised attenders with an aneurysm of any size. These are both assumed to be the same as the proportion observed in the attenders for the baseline model, but these parameters

enable investigation of the impact on results if these proportions were higher or lower in non-attenders or non-visualised attenders, assuming the distribution of sizes remains constant.

## 2.2.4 Model cost parameters

In addition to the parameters defining the pathway of individuals through the model, there is a set of parameters that determine the costs accrued as the model progresses. Costs are assigned to each costed event modelled: invitations, scans, consultations and operations. Parameters relating to the accrual of costs are defined in Table 2.3.

Parameter name	Description
<i>Event costs</i>	
inv_cost	Cost of initial invitation to screening
reinv_cost	Cost of re-inviting non-attenders
scan_cost	Cost of first scan
recall_cost	Cost of recall scan for those with detected AAA
consult_cost	Cost of consultation for considering elective surgery
emer_cost	Cost of emergency operation
elec_cost	Cost of elective operation
scans_grp2*	Number of scans per cycle for those with detected small AAA
scans_grp3*	Number of scans per cycle for those with detected medium AAA
consults*	Mean number of consultations per elective operation undertaken
<i>Discounting</i>	
disc_eff	annual discount rate for outcomes
disc_cost	annual discount rate for costs

\* Used in calculation of other parameters only; not direct parameters in model

Table 2.3 Parameters relating to model costs

## 2.3 Assumptions and simplifications: specific components

There are a number of assumptions and simplifications that have been made with respect to various aspects of this model of AAA screening, some of which have been briefly mentioned already. The assumptions regarding the model structure and interpretation of resulting parameters (as opposed to the estimation of the parameters) will now be discussed in detail. Those relating to specific components in the model are included in this section (2.3), whilst those relating more generally to the whole model can be found in Section 2.4.

### 2.3.1 AAA size and growth

Size groupings (no AAA, small, medium and large AAA) have been chosen to reflect the screening recall strategy employed in the MASS trial. However, the cut-off points between the categories and defining that an AAA ( $\geq 3.0$ cm) are somewhat arbitrary, and other screening programmes have used different cut-offs. The recall intervals employed for different size categories also vary between screening programmes. This model uses a three-month cycle length to reflect the three-month recall scans for medium AAAs in MASS, and therefore makes the assumption that the outcomes and costs for a three-month cycle can be estimated from observations over a longer period (e.g. over one year for recall scans for small aneurysms). This also means the changes between size states due to growth and loss to recall follow-up are modelled more smoothly over time than the results of annual observations in the trial. The model allows only for transitions between increasing consecutive AAA size states, although apparent decreases are possible in practice as a result of the measurement error associated with ultrasound scans. This will result in an overall overestimation of the numbers of individuals in the larger AAA size states.

Assumptions must also be made about the distribution of aneurysm sizes amongst non-attenders and non-visualised attenders. In the model, the distributions are assumed to be the same as that observed among attenders; however, since non-attenders are known to be a less fit group, it is possible that there is a higher proportion of individuals with an aneurysm in the non-attending group. Ultrasound screening results in non-visualisation of the aorta in around 1% of individuals attending for a scan. It has been suggested that this group are likely to comprise a higher proportion of individuals with smaller aortas, as these tend to be more difficult to visualise than large or aneurysmal aortas. There is currently no published evidence regarding the comparative mortality in this group, although the MASS data indicates it may be higher than those with a visualised aorta, particularly when compared to those with an aorta identified as non-aneurysmal. However, in the model it is assumed that the progress of individuals with a non-visualised aorta can be estimated by assuming the same proportion with an aneurysm, and the same distribution across sizes, as observed in visualised attenders.

### **2.3.2 Consultation for considering elective surgery**

The decision at the consultation for considering elective surgery has been modelled by a three-outcome node comprising returning to screening, contraindication/ patient refusal, and proceeding to elective surgery. The number of consultations per patient is estimated by the mean number recorded per patient entering the large detected state in MASS (excluding those dying from non-AAA or rupturing during follow-up). This may be an underestimate since a maximum of two consultations per patient could be recorded in the trial. However, this value is also applied to patients initially returned to the medium size state each time they re-enter the large AAA state in the model. The consultation and any resulting elective surgery are modelled in two consecutive three-month cycles in an effort to account for some of the delays observed in practice in receiving a consultation and between the appointment and surgery. However, the use of the ‘elective surgery pending’ state is specific to the three-month cycle length, and it is also unclear how accurately this will reflect delays in the long-term.

For those contraindicated for or refusing elective surgery following the consultation, the assumption is made that they cannot then later undergo elective surgery. In practice, some patients may be reconsidered for elective surgery even if unfit at the first consultation. These additional elective operations may be accounted for in the model by proportionally increasing the relevant parameter in the large detected AAA pathway; this is particularly important for costs, owing to the significant cost of each elective surgery. Patients returned to screening at the consultation are assumed to be returned to the medium detected state. Whilst in reality it would be possible to receive a measurement corresponding to a small AAA, this is a rare occurrence, since ultrasound measurement error accounts for most of those returned to screening. Since these patients are returned to the medium detected state, it is assumed that the parameters for this state reasonably represent their progress, although they are likely to have larger AAAs than the mean size in the medium size state.

### 2.3.3 Elective and emergency surgery

Elective and emergency operations carry considerably higher costs than other events in the screening model, so the assumptions surrounding the numbers proceeding to these events are perhaps the most crucial. This model specifies consultation for considering elective surgery after reaching a single criterion, the large size state. This is defined as those  $\geq 5.5\text{cm}$ , as used in MASS, but there would be an impact on outcomes and costs if this cut-off was defined as  $5.0\text{cm}$  or  $6.0\text{cm}$ , as has been employed in other screening programmes. In MASS, an additional criterion for considering elective surgery was implemented, using a “rapid expansion” of  $\geq 1.0\text{cm}$  in a one-year period to initiate the consultation process. This meant that in practice, individuals with rapidly expanding aortas with a diameter  $< 5.5\text{cm}$  could still be referred for elective surgery. However, few of those receiving a consultation actually underwent elective surgery on this criterion alone in the trial. Recent guidelines also recommend the use of a size criterion only, hence this model provides a realistic representation of any future AAA screening programme in only including the possibility of proceeding to elective surgery after reaching the large size state. Excluding modelling of individuals with a rapidly expanding aorta through the decision process, however, does not represent the MASS trial strategy well. So although this model structure may provide more accurate estimates for a national screening programme, it may not provide good estimates of the outcomes and costs observed at four years follow-up in MASS. For those who do undergo consultation in the model, costs are applied for an expected mean number of consultations, determined by an input parameter (“consults”), since many patients receive more than one whether or not they ultimately undergo an elective operation.

Following successful elective or emergency surgery, the individual passes into one of two “surgery survivor” states (depending on whether the operation was elective or emergency), and remains in this state for all subsequent cycles until death. This mortality is only included in the model as non-AAA related mortality; it is assumed that AAA-related mortality cannot occur  $> 30$  days post-operation. However, a small number of individuals may die of AAA-related causes  $> 30$  days after an AAA repair, often as a

result of late complications from the operation, and due to the arbitrary nature of the cut-off of 30 days for post-operative mortality.

#### **2.3.4 Incidental detection**

It is assumed that no incidental detection has taken place in the population prior to the start of the model, hence for the no-screening arm, it is assumed that no individuals start the model in a detected state. This is a reasonable assumption in modelling the observed four-years results of the MASS trial, since those with known AAAs were excluded prior to randomisation. However, it may not so accurately reflect the situation for a national screening programme, where some incidental screening may have taken place prior to the age at which a invitation is issued for systematic screening.

The probability of incidental detection of a previously undetected aneurysm is assumed not to depend on size, although the use of three separate parameters containing the same estimate for the baseline model (opp\_grp2, opp\_grp3, opp\_grp4) facilitates investigation of the effect of these estimates being different. It is also assumed that incidental detection of an aneurysm has no influence on parameters relating to detected AAA states. Decision pathways for detected AAA states use a single estimate of parameters for both aneurysms detected through systematic screening and those detected via incidental screening. In reality, individuals with an aneurysm detected incidentally are more likely to have co-morbidities, and consequently may have a higher all-cause mortality, may be more likely to be declared unfit for elective surgery, and may have a lower probability of a successful operation.

Costs of incidental detection are applied in the model for long-term cost-effectiveness only, since surveillance following incidental detection (and therefore also the costs associated with surveillance events) was not recorded in MASS. In the long-term model, recall scan costs are applied to AAAs detected incidentally in both the screened and unscreened arms, excluding the cost of the scan at which the aneurysm was initially identified, since this identification resulted from another procedure.

### **2.3.5 Non-AAA mortality**

Deaths not related to AAA are modelled by application of either national death rates (long-term model) or rates observed in the MASS trial (four-year model). These rates are estimated separately for each three-month cycle of the model, and hence allow for the increasing age of the population as the model progresses. In each cycle, a single parameter estimate for the probability of non-AAA-related mortality is applied across all states. However, this overall estimate is likely to underestimate mortality in non-attenders; in MASS, this group was noted to have significantly higher all-cause mortality compared to attenders. Furthermore, since the proportion of non-attenders with an aneurysm is assumed to be the same as the proportion observed in the attenders, underestimating mortality in the non-attenders will cause the model to predict too many aneurysm growth transitions and too many ruptures in this group.

### **2.3.6 Costs**

In the current proposed model, health-related costs other than for AAA surgical procedures (excluding secondary procedures of this nature) are not accounted for. However, the inclusion of such costs may be important for long-term modelling since AAA screening is known to result in increased life-years, thus enabling more time for other non-AAA health problems and their associated costs in the invited group. Costs relating to non-AAA health problems would be difficult to estimate and incorporate in the model, although in the absence of their inclusion the model is likely to underestimate the difference in costs between the control and invited arms.

## **2.4 Assumptions and simplifications: general**

### **2.4.1 Markov modelling assumptions**

Some of the major assumptions in the model result from the use of a Markov structure to model the screening programme. The Markov model cycles through transition states by

means of decision pathways, and by nature makes the assumption that all individuals starting a cycle in a given state can be modelled in the same way. The implication is that past events and the route to arriving in this state have no influence on subsequent parameters in the decision pathway, i.e.

$$P(X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i) = P(X_n = j | X_{n-1} = i)$$

where  $X_n$  is a discrete-time stochastic process occurring at time  $n$  ( $n=0,1,2,\dots$ ), and  $i, j$  refer to states in the Markov chain. This assumption is commonly described as the Markov property, and may in fact be incorrect for many parts of this model; for example, the individuals returned to the medium detected state following a consultation may have different transition probabilities than other individuals in this state. It is also implied that the length of time spent in a given state by a particular individual has no influence on transition parameters. Furthermore, this model applies time-homogeneous parameters (apart from non-AAA death rates, which change for each year of age, and hence each cycle of the model), although many parameters may depend on age (i.e. time) and other factors. The assumption of time-independent parameters (the stationarity assumption) occurs when for  $n \geq 1$  and all  $i, j$ :

$$P(X_n = j | X_{n-1} = i) = p_{ij}$$

where  $p_{ij}$  defines a transition probability for movement from state  $i$  to state  $j$  for all values of  $n$ . Despite these fundamental assumptions, the Markov structure provides a useful simplified mechanism for modelling a complex process.

## 2.4.2 Censoring

Due to staggered entry times into the MASS trial over a period of two years, the observed follow-up times are censored between years three and five, giving an overall mean follow-up of approximately four years. The Markov model is to be used principally to estimate both lifetime costs and outcomes, but will additionally be run over a four-year period to check the results against the observations from the trial. In this latter version of

the model, censoring must be implemented in order to obtain comparable estimates of numbers and timing of events. This is achieved in the model by censoring in each three-month period the observed proportion censored in MASS. The numbers censored are removed proportionately from all states of the model, and a half-cycle correction is applied to the timing of these removals. This correction is carried out alongside a similar correction for the timing of deaths; the whole procedure is discussed in detail in Section 2.4.3.

### **2.4.3 Timing of outcomes and costs**

Owing to the use of a three-month cycle length, costs incurred over a longer period of time in the screening programme must be assumed to occur equally over a number of three-month periods. This is the case for recall scans for individuals with a small aneurysm, which take place annually, although the cost for such a scan must be spread equally over four three-month cycles. This will result in some inaccuracy in the costs estimated for recall scans in this group, as individuals may move into other states during a one-year period, and yet have incurred recall scan costs corresponding to a fraction of an ultrasound scan.

The method of assigning both costs and outcomes for each cycle of the model represents further assumptions regarding the model as a whole. Costs are attached to each costed event occurring within a cycle, and are discounted according to the time elapsed since the start of the model. For the initial screening, an assumption is made that this takes place instantaneously, to enable immediate population of detected and undetected starting states in the invited arm. This implies the accrual of screening costs prior to the start of cycle 1.

For each cycle, a half-cycle correction is applied to entry to absorbing states (and to censoring when applied in the four-year model, see Section 2.4.2), to enable more accurate estimation of life-years accrued per cycle. This is necessary because any movement to a different state is assumed to occur at the end of each cycle in the model; in practice these transitions will occur throughout the three-month period. This imprecision in timing includes transitions to the death states, creating inaccuracies in the

estimation of life-years for each arm. The half-cycle correction instead assumes that transitions into the death states occur exactly half-way through the cycle, reflecting the expected average timing of a transition. The practical implementation of these corrections involve accounting for half of the deaths and censoring events at the start of a cycle and half at the end, with other transitions calculated for those at risk after the first half are removed from the risk-set:

$$y_{ni} = x_{ni} - \frac{x_{ni}c_n}{2} - \frac{d_n}{2} \left( x_{ni} - \frac{x_{ni}c_n}{2} \right)$$

where  $y_{ni}$  is the number of individuals in state  $i$  in cycle  $n$  after the first half of non-AAA deaths and censoring events are removed (it is to  $y_{ni}$  that transition probabilities are applied),  $x_{ni}$  is the number of individuals in state  $i$  at the start of cycle  $n$ ,  $d_n$  is the probability of non-AAA death in cycle  $n$ ,  $c_n$  is the probability of censoring in cycle  $n$ .

The number of individuals in state  $i$  at the end of cycle  $n$  (i.e. the number in state  $i$  at the start of cycle  $n+1$ ) comprises the number remaining in the state ( $a_{ni}$ ) and the number entering the state ( $b_{ni}$ ):

$$a_{ni} = \left[ y_{ni} - y_{ni} \mathbf{p}_{ni} - \frac{\frac{d_n}{2} \left( x_{ni} - \frac{x_{ni}c_n}{2} \right) (x_{ni} - x_{ni} \mathbf{p}_{ni})}{x_{ni} - x_{ni} \mathbf{q}_{ni}} \right] \left[ \frac{x_{ni} - x_{ni}c_n}{x_{ni} - \frac{x_{ni}c_j}{2}} \right]$$

$$= y_{ni} (1 - \mathbf{p}_{ni}) \left[ 1 - \frac{\frac{d_n x_{ni}}{2} \left( 1 - \frac{c_n}{2} \right)}{y_{ni} (1 - \mathbf{q}_{ni})} \right] \left[ \frac{1 - c_{ni}}{1 - \left( \frac{c_{ni}}{2} \right)} \right]$$

where  $\mathbf{p}_{ni}$  is a vector of transition probabilities ( $p_{ni}$ , the sum of probabilities from a number of decision nodes in the pathway) resulting in exit to other states (excluding by censoring or non-AAA death), and  $\mathbf{q}_{ni}$  is a vector of transition probabilities resulting in

AAA death (since the second half of non-AAA deaths must only be applied to those not dead by AAA already). Censoring of those exiting to other states occurs on entry to these other states (see calculation of  $b_{ni}$ ).

$$\begin{aligned} \sum_k b_{nik} &= \left[ \frac{y_{nk} p_{nk} - \frac{y_{nk} p_{nk} d_n}{2} \left( x_{nk} - \frac{x_{nk} c_n}{2} \right)}{y_{nk} - y_{nk} \mathbf{q}_{nk}} \right] \left[ \frac{y_{nk} - y_{nk} c_n}{y_{nk} - \frac{y_{nk}}{2}} \right] \\ &= p_{nk} \left[ \frac{1 - \frac{x_{nk} d_n}{2} \left( 1 - \frac{c_n}{2} \right)}{1 - \mathbf{q}_{nk}} \right] \left[ \frac{1 - c_n}{1 - \frac{c_n}{2}} \right] \end{aligned}$$

where  $k$  is some state other than  $i$  ( $i \neq k$ ), and  $p_{nk}$  is the sum of transition probabilities leading from state  $k$  to state  $i$  in cycle  $n$ .

#### 2.4.4 Discounting

In accordance with current guidelines and general practice, both costs and outcomes are discounted over time. This enables time preferences for delaying costs to be incorporated and valid conclusions drawn concerning cost-effectiveness by similarly discounting outcomes. Discounting is applied by:

$$V_t = V_0(1+r)^{-t}$$

where  $V_t$  is the discounted value at time  $t$ ,  $V_0$  is the equivalent value at time 0, and  $r$  is the discount rate. At the time of the principal MASS publications (2002), the recommended discount rates for UK trials were 6% for costs and 1.5% for outcomes. These values are therefore used in the model of four-year cost-effectiveness for comparison with the MASS trial results. However, current recommendations suggest using 3.5% for both costs

and outcomes, enabling easier interpretation of long-term discounted cost-effectiveness estimates; these values are applied to the long-term model.

#### **2.4.5 Sensitivity and specificity of the screening test**

The assumptions made concerning individuals with a non-visualised aorta, and the potential for misclassifications of AAA sizes have already been discussed. This must also be brought together in consideration of the sensitivity and specificity of ultrasound as a screening test. Since there is measurement error associated with the scan, and some individuals cannot be visualised, there is potential for both false positive and false negative outcomes at the initial scan. The basic model structure does not incorporate this, but this aspect of the screening could be investigated using a small number of additional parameters to modify the distribution across starting states.

#### **2.4.6 Parameter dependence on covariates**

All parameters are assumed to be the same for all individuals simulated by the model. However, it is likely that at least some of the parameters may be dependent on individual covariates such as age.

### 3 Estimation of parameters

A large number of parameters have been described as components for this model for AAA screening. The majority of these can be estimated from the MASS trial data, although it may be possible to increase the accuracy of these estimates by incorporating a systematic review of the literature with regard to these parameters. A small number of the model parameters must be estimated in this way however, where the MASS data does not contain the relevant information. This is necessary for the estimation of aneurysm growth in those initially scanned as non-aneurysmal and the rate of incidental detection in the uninvited and non-attending populations.

The transition parameters to be estimated for the model are required in the form of the probability of an event over the time-period of a model cycle. However, whether from trial or systematic review data, the relevant information is frequently available over a longer time-period or as a rate. In these situations, the rate can be converted into a probability over the required time-period:

$$P(3 \text{ mons}) = 1 - \exp^{-\lambda}$$

Where  $\lambda$  is the rate per person-3-months, equivalent to the rate per person-year / 4 since rates and probabilities are assumed to be constant over time. Probabilities of an event over a longer period can be converted to the necessary length of time:

$$P(3\text{-mons}) = 1 - (1 - P_t)^{1/4t}$$

Where  $P_t$  is the probability of an event over  $t$  years.

The use of the MASS data to estimate the remainder of the parameters is described first, followed by a description of parameters estimated from systematic review, and a discussion of the assumptions involved in applying these parameter estimates to the model.

### **3.1 Estimation of individual model components from MASS**

The use of the MASS four-year data to estimate parameters for use in the long-term model of the cost-effectiveness of AAA screening makes a number of broad assumptions about the generalisability of estimates from this source. It is assumed that these estimates obtained from a starting population of 65-74-year-old men from areas known to be less socially deprived than the median for England and Wales can produce interpretable results in a model for 65-year-old men. It is further assumed that all these parameters are time homogeneous (with the exception of non-AAA death rates which vary by cycle, and thus by age). Some of the resulting parameter estimates are based on very few events in the MASS dataset; the large uncertainty surrounding such parameters is not addressed in the current report.

#### **3.1.1 Calculation of person-years used in parameter estimations**

A number of three-month probability parameters in the model are based on rates of events in MASS (such as ruptures), while others are based on probabilities of outcomes in MASS (such as 30-day mortality following elective surgery). For those based on rates, person-years must be calculated and where necessary, separated into person-years in each AAA size group. This requires a number of assumptions to be made about the entry and exit of individuals into size groups, which are now discussed.

Although AAA size decreases may be recorded, it is thought the majority of these apparent decreases are due to measurement error in ultrasound readings. Furthermore, management of the AAA is not altered when a decrease is recorded; once the 4.5 cm threshold for the medium size is met, the individual remains in the three-monthly recall cycle regardless of any subsequent scan of less than 4.5 cm. In the calculation of person-years therefore, the last known maximum measurement is used in determining the size group to which person-years are contributed. The date of randomisation is used to initiate entry into the first known size group, as recorded at the initial scan. Entry into subsequent size groups is initiated on the date of the relevant scan, and exit from the previous size group assumed to take place a day before this scan took place. Individuals not visualised

at the initial scan or not attending screening do not contribute person-years to any size group; person-years accrued in the non-aneurysmal size group are taken only from those with a known  $<3.0\text{cm}$  initial scan. Where individuals are lost to recall follow-up, contribution to person-years is assumed to terminate on the last known scan date.

Accrual of person-years following a consultation for considering elective surgery is determined by the outcome of the consultation. For individuals returned to screening (defined as not referred to/by GP, not AAA at consultation, or return to screening decision), person-years are counted to the medium size state from the date of the consultation if the scan measurement corresponding to the consultation is  $<5.5\text{ cm}$ . However, if the individual is returned to screening at consultation despite the corresponding measurement being  $\geq 5.5\text{ cm}$ , person-years are counted to the large size state from the date of consultation. Decisions to return to screening following a referral for expansion (growth  $\geq 1.0\text{cm}$  in a year) only are ignored for the purposes of counting person-years and consultation events. For those contraindicated for elective surgery at consultation (defined as patient refusal of consultation or operation, or declared unfit for elective surgery), person-years are counted to the contraindicated state from the date of consultation. For all other outcomes of consultation (for example; referred on, social deferral, elective operation planned), the individual is assumed to continue accruing person-years in the large state until an AAA operation, rupture or death takes place. These events similarly indicate exit from all of the size states. No person-years or events are counted for individuals with an AAA operation before randomisation, but not identified in the pre-randomisation exclusions (these individuals remain in the analysis for the purpose of intention-to-treat analyses). The actual person-years contributed to each size state in MASS based on the assumptions discussed are given in Table 3.1.

### **3.1.2 Parameters defining distribution across starting states**

Individuals entering the model are initially distributed across a number of starting states. These states relate to the detection status and size of any aneurysm, and are described in section 2.2.1. For the invited to screening arm, it is assumed that the proportion of the group attending the initial scan, and the proportion of these with an aneurysm detected is

the same as was observed in MASS. Amongst those with an aneurysm detected, the distribution across size groups is also taken directly from the MASS data. Furthermore, this same proportion with an aneurysm and distribution of sizes is assumed to apply to the non-attenders and non-visualised attenders. It is not clear however, if either of these groups may include a higher proportion with an aneurysm, and if aneurysms in these groups are likely to be larger. Similarly, the proportion of the uninvited group as a whole with an aneurysm is assumed to be the same as this proportion among attenders in MASS. The proportion of individuals with an incidentally detected aneurysm at the start of the model (i.e. incidentally detected before the age at which the model is started) is assumed to be zero for both the invited and uninvited to screening arms.

The parameters relating to the distribution of individuals across starting states that are obtained directly from MASS are given in Table 3.1; the remaining parameters required for the starting states are derived from these.

<b>Starting state parameter</b>	<b>Proportion in MASS</b>	
attend	27147 / 33839	= 0.80
nonvis	329 / 27147	= 0.012
s_grp2det	944 / 27147 x attend	= 0.028
s_grp3det	223 / 27147 x attend	= 0.0066
s_grp4det	166 / 27147 x attend	= 0.0049

Table 3.1 Parameter estimates relating to model starting states

### 3.1.3 Rupture rates

Rupture rates are estimated separately for each aneurysm size group (small, medium, large) detected in MASS. It is then assumed that these rates can be applied to both detected and undetected aneurysm states for small and medium aneurysms. Since the invited arm in MASS received only surveillance at these sizes, it is a reasonable assumption that the rupture rates in these detected states should be similar to rupture rates for the same size groups undetected. However, the surgical intervention in the large detected state means that ruptures are considerably less in this group than would be expected in those with a large, undetected aneurysm with no elective intervention. The difference arises because the intervention in the invited arm results in a different range of

sizes within the large size group (there is no upper limit to the aortic diameter in this group). Most detected aneurysms will be operated on close to the criterion for elective surgery (in this model, 5.5cm), hence there will be far less very large aneurysms in the large detected group compared to the large undetected group, resulting in different probabilities of rupture. Although the rupture rate for the large undetected aneurysm state cannot be estimated directly from MASS, because this parameter relates to an undetected state, there is no available evidence elsewhere to use in the estimation. Consequently, it is assumed that the rate lies somewhere between the rupture rate for large detected AAAs and that for those contraindicated for elective surgery (in whom the aneurysm continues to grow, but who are selected less fit group). Validation against the MASS trial in terms of numbers of ruptures enables the parameter to be estimated within this range.

<b>AAA state</b>	<b>Person-years</b>	<b>Growth to next size state*</b>	<b>Ruptures*</b>
Non-aneurysmal	100960	not available in MASS	0.000015 (6)
Small AAA	2618	0.025 (257)	0.000095 (1)
Medium AAA			
Before consultation	713	0.087	0.0023
After returned from consultation	58	(269)	(7)
<i>Total</i>	<i>771</i>		
Large AAA			
Before consultation	166	-	0.016
After returned from consultation	55		(14)
<i>Total</i>	<i>221</i>		
Contraindicated for elective surgery	64	-	0.035 (9)

\* Rate / person-three-months (no. events in MASS)

Table 3.2 Parameter estimates relating to model growth and rupture transitions

In the model, all individuals without an aneurysm (detected or undetected) are grouped together into one, “no AAA” state. In the invited arm, this includes those with a non-aneurysmal initial scan, those not visualised at the initial scan, and non-attenders, and the same proportion of the population is assumed to be in this starting state in the uninvited arm. The rupture rate for this state can also be estimated directly from MASS by

combining data for all these subgroups. The observed data in MASS relating to the estimation of rupture parameters is given in Table 3.2.

### 3.1.4 Dropout rates

At each recall scan for those with a detected aneurysm, there were a number of non-attenders, with a total of 230 from the small size group, 62 from the medium and 10 from the large size group lost to follow-up over the four years. This has been incorporated into the model by returning these individuals to the relevant undetected size state. The probability of this dropout occurring can be estimated from MASS, but whether this probability varies over the course of the follow-up, and whether it depends on AAA size must also be determined. This was investigated using Poisson regression for the number of drop-out events per three-month period. Pooled data for all individuals is used because a date of drop-out is not recorded; individuals are instead described as dropping out of follow-up in the period in which they attended their last known scan. Drop-outs from the small AAA size group are additionally combined over each set of four three-month periods, reflecting the annual recall scan strategy for this group. Those with large AAAs were not included in the analysis, since only six individuals with a large AAA were lost to recall follow-up (excluding any not continuing surveillance after contraindication for elective surgery), of whom four had moved to another screening programme.

	<b>Effect of 3-month period</b>		<b>Effect of AAA size</b>	
	IRR (95% CI)	p>  z	IRR (95% CI)	p>  z
<i>Poisson regression</i>				
Period only	1.05 (1.03, 1.08)	<0.0005	-	-
Period + size	1.05 (1.03, 1.08)	<0.0005	0.80 (0.60, 1.06)	0.12
<i>Negative binomial regression</i>				
Period only	1.06 (0.01, 1.11)	0.02	-	-
Period + size	0.10 (0.01, 1.11)	0.02	0.71 (0.40, 1.26)	0.24

Table 3.3 Analysis of association of drop-out rates with time and AAA size

Poisson regression of loss to recall follow-up was carried out on 27 data points (5 relating to small AAAs, 22 to medium AAAs). A model adjusted for three-month period and AAA size group (small v medium) suggested overdispersion ( $\chi^2$  goodness-of-fit test = 86.4,  $p > \chi^2_{24} < 0.0005$ ), and the data was subsequently reanalysed using negative binomial

regression. This provided no evidence of size group predicting the drop-out rate, but some evidence of an association of time with drop-out (Table 3.3).

For simplicity, a single, time-independent value will be used for the probability of drop-out in the main model, although this may underestimate dropout rates in a Markov model running over a longer time period. A time-varying parameter will be assessed in the sensitivity analysis of the model. The estimate for the constant parameter is taken from a negative binomial model with no predictor variables:

$$y_i \sim \text{Poisson}(\mu_i^*)$$

$$\mu_i^* = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \ln(\text{exposure})_i + u_i)$$

$$e^{u_i} \sim \text{Gamma}(1/\alpha, 1/\alpha)$$

where  $\beta_0 = -3.70$ ,  $(\beta_1 x_{i1} + \dots + \beta_k x_{ik}) = 0$  (since no predictor variables included),  $e^{u_i} = 1$ , and exposure = 1 person-year, giving a rate estimate of 0.025, then applied to the model.

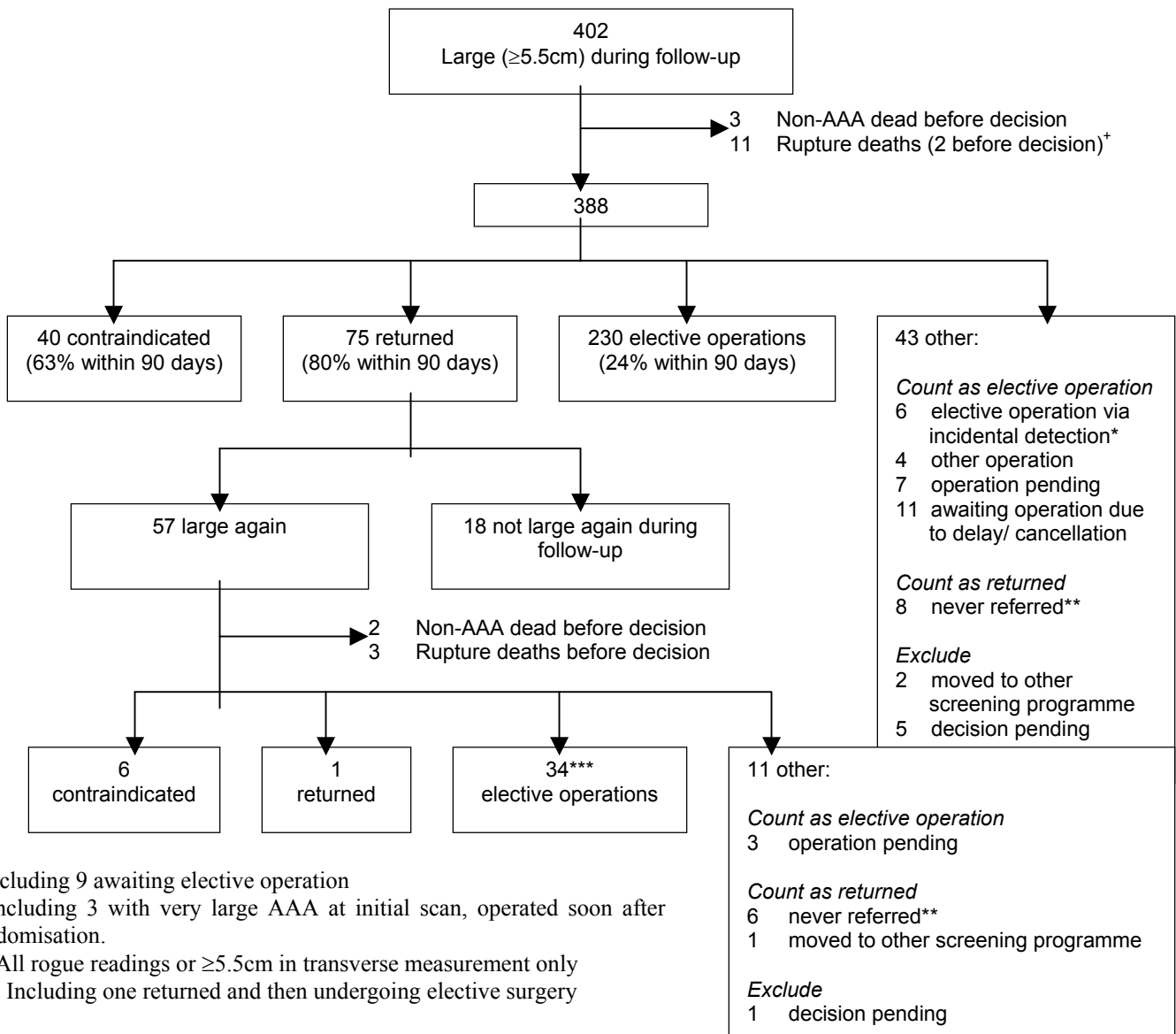
### 3.1.5 Aneurysm growth parameters

The growth of detected aneurysms from small to medium and from medium to large size groups can be estimated directly from MASS, using data from recall scans to identify the number of transitions taking place between the groups over the observed period (details given in Table 3.2). These estimates are then reasonably assumed to apply also to the undetected states, since there is only surveillance at this stage of follow-up for those in the detected states. Development of an aneurysm in those initially considered to be non-aneurysmal cannot be estimated from MASS, since there was no follow-up in the trial of those with a diameter of <3.0cm at the initial scan, those not visualised or non-attenders. Estimation of this parameter from external data is discussed in Section 3.2.1.

### 3.1.6 Estimation of outcomes following referral

Perhaps the most difficult node to estimate parameters for is that relating to outcomes from consultation. Since the large detected AAA state is a ‘tunnel’ state in the model, the

parameters for this node must represent outcomes ever” rather than estimation of probabilities from transition rates over a period of time. There is also difficulty in incorporating data from observed individuals in MASS with event combinations not permitted in the model, such as receiving a decision to proceed with elective surgery, but rupturing before the procedure. This has implications for both effects and costs in the model. The event combinations actually observed in MASS are described in detail below (Fig 3.1), together with a description of how these data are used to produce the parameter estimates used in the model.



<sup>+</sup> Including 9 awaiting elective operation

\* Including 3 with very large AAA at initial scan, operated soon after randomisation.

\*\* All rogue readings or ≥5.5cm in transverse measurement only

\*\*\* Including one returned and then undergoing elective surgery

Fig 3.1 Progression of MASS patients reaching 5.5cm during four-year follow-up.

... decisions for those entering the large state for a second time (i.e. those returned to screening, but referred for reaching 5.5cm

again) should be included. Individuals dying (either non-AAA or fatal ruptures) before a decision was made regarding elective surgery should be excluded, since these events are included in the model for large AAAs before the decision process. Additionally, the small number of individuals for whom the decision is unknown should be excluded, equivalent in terms of the parameter estimation of assuming this group have the same distribution of decisions as those with a known decision. The total number of events (AAAs reaching 5.5cm) contributing to the calculation of the decision outcome parameters is therefore 431. The estimation of the parameters relating to each of the decision outcomes included in the model is given in Table 3.4.

<b>Decision if large AAA and reach decision node</b>	<b>No. decisions in MASS in follow-up period (n=431)</b>	<b>Probability of decision</b>
Contraindicated/ refused elective surgery	46	0.107
Returned to screening	90	0.209
Elective surgery	295	0.684

Table 3.4 Parameter estimates relating to decision at consultation

Costs are attached to the consultation process by multiplying the cost of each modelled large AAA reaching consultation by the mean number of consultations for patients with a large screen-detected AAA in MASS. Based on recorded consultation dates for the 402 individuals described in Fig 3.1, 226 had two consultations (this includes those returned to screening reaching large again in addition to a variety of reasons, such as undergoing fitness tests), 174 had one consultation, and two had no consultations recorded (one underwent another operation type a few days after reaching large, and one was never referred due to only meeting the size criterion in the transverse measurement). This gives a mean of 1.6 consultations per patient reaching the large AAA size, although this is likely to be an underestimate since a maximum of only two consultations could be recorded in MASS. Furthermore, this is also applied to costs for those reaching large size through incidental detection; it is possible that the mean number of consultations is different for this group compared to screen-detected aneurysms.

### 3.1.7 Emergency surgery parameters

Both the probability of having emergency surgery once ruptured and the probability of survival (>30 days) following an emergency operation can be estimated from MASS. Whilst the data from the invited group would enable each of these parameters to be estimated separately for each aneurysm size group (including the “no aneurysm” state), the data from the control group could not be used in such estimations. Furthermore, it is conceivable that the probability of reaching hospital and receiving emergency surgery and the probability of a successful operation would not differ according to aneurysm size. This assumption cannot be tested very satisfactorily owing to the small number of ruptures and emergency operations in the invited arm (see Table 3.5). However, using data from 37 ruptures in this arm, Fishers exact test for a difference in the proportion of ruptures receiving emergency surgery across five Markov states (no aneurysm, small AAA, medium AAA, large AAA, contraindicated for elective surgery) gave a p-value of 0.05, providing borderline evidence that the probability of emergency surgery following rupture is different in the groups. For simplicity, a pooled estimate will still be employed in the principal model; this assumption can be tested in sensitivity analyses. A similar test for probability of a successful emergency operation across the same five groups (this time based on only 20 emergency operations in the invited attenders) gave a p-value of 0.07, and hence the same conclusion of borderline evidence of a difference across groups, with a simple pooled parameter to be applied in the model and investigated later.

	<b>No. ruptures</b>	<b>No. emergency operations if rupture</b>	<b>No. post-operative deaths if emergency operation</b>
Controls	138	64 (46%)	26 (41%)
Invited: attenders			
No AAA	6	2 (33%)	1 (50%)
Small AAA	1	1 (100%)	0 (0%)
Medium AAA	7	4 (57%)	0 (0%)
Large AAA	14	11 (79%)	2 (18%)
Contraindicated	9	2 (22%)	2 (100%)
Lost to recall follow-up	4	1 (25%)	1 (100%)
Invited: non-attenders, non-visualised attenders	25	7 (28%)	3 (43%)
<b>Total</b>	<b>204</b>	<b>92 (45%)</b>	<b>35 (38%)</b>

Table 3.5 Observed events in MASS relating to emergency operations

The pooling of data from both the control and invited arms will importantly decrease the uncertainty surrounding the parameter estimates, which would be considerable if size

groups with very little observed data had each required a separate estimate. The final pooled data give an estimate of  $90 / 204 = 0.44$  for probability of receiving emergency surgery given rupture, and  $32 / 90 = 0.36$  for death following emergency surgery.

### 3.1.8 Post-operative death rates

Death within 30 days of elective surgery is estimated separately for operations via incidental detection and via screen-detection). This is significant because aneurysms detected incidentally may be detected at an older age, without monitoring may have become very large, and are more likely to be in individuals with co-morbidities which led to the scan taking place. These factors suggest that the post-operative mortality following elective surgery on an incidentally detected AAA may be higher than for a screen-detected AAA; this is confirmed by post-operative mortalities observed in MASS. For operations via systematic screening, the proportion dead within 30 days was 11/295 (3.7%), whereas for those with an incidentally detected AAA, this proportion was 13/131 (9.9%). The data from MASS on which these estimates are based is summarised in Table 3.6.

	No. elective operations (no. dead within 30 days)	
	Systematic screening	Incidental screening
Controls	-	100 (9)
Invited	295 (11)	31 (4)
Total	295 (11)	131 (13)

Table 3.6 Observed events in MASS relating to elective operations

### 3.1.9 Non-AAA death rates

Non-AAA death rates are estimated from difference sources depending on the purpose of the model. For the long-term model projections, national death rates are used for each year of age, with the corresponding national rates for AAA subtracted out. For the four-year model to be validated against the four-year results of the MASS trial, non-AAA

death rates are modelled using death rates from each three-month period of the trial. This is necessary because individuals randomised into MASS are known to be less socially deprived than the median for England and Wales, and hence have different death rates from the national average (see Table 3.7). Furthermore, national death rates would have to be applied for the mean age in MASS at the start of the trial and then assume this mean increases by one year for each year of the model, although in practice it will increase by slightly less than one year for each year of follow-up in MASS.

<b>Approximate mean age in MASS</b>	<b>National non-AAA annual death rate</b>	<b>MASS-specific non-AAA annual death rate (no. deaths based on)</b>
69	0.027	0.022 (1511)
70	0.030	0.027 (1799)
71	0.034	0.029 (1874)
72	0.037	0.031 (1594)

Table 3.7 National and MASS-specific death rates

Non-AAA deaths are estimated separately for those in the contraindicated for elective surgery state, since this group will be less fit than the norm. In MASS, 15 events are observed in 64 person-years at risk, translating to a 3-month probability of 0.057.

### 3.1.10 Cost parameters

Fixed cost parameters are estimated in the model directly from the data collected for the evaluation of the four-year cost-effectiveness of the MASS trial, calculated for the financial year 2000-1. These costs are all based on combined data from all four centres, and are given in Table 3.8. Costs for emergency and elective operations are based on 217 and 361 costed events respectively. For the long-term extrapolation of the model, cost and cost-effectiveness estimates are required in real terms, and hence are inflated to 2004-5 prices, also given in Table 3.8. This involves multiplying by the estimated Hospital and Community Health Services (HCHS) pay and prices inflation index for each of the years 2001-2, 2002-3, 2003-4, and 2004-5 (5.1%, 3.6%, 5.5%, and 3.8% respectively).

<b>Event</b>	<b>Cost (2000-1)</b>	<b>Cost (2004-5)</b>
<i>Screening costs</i>		

Invitation	£1.31	£1.56
Re-invitation	£1.28	£1.53
Initial scan	£19.08	£22.75
Recall scan	£46.04	£54.90
<i>Surgery costs</i>		
Consultation	£309.88	£369.49
Elective operation	£6908.75	£8237.81
Emergency operation	£11175.63	£13325.52

Table 3.8 Original and inflated costs for costed events in model

## 3.2 Estimation of components from non-MASS sources

### 3.2.1 Growth in “no aneurysm” state

In the proposed model, the proportion of attenders with a non-aneurysmal or non-visualised initial scan begin in a “no AAA” state. The progress of this group is then modelled in terms of aneurysm growth, detection and rupture rates over time. However, in MASS this group of individuals received no further scans, and only information regarding operations, rupture and death was collected. The development and growth of aneurysms in this group must be obtained from elsewhere.

A limited systematic review of the literature was carried out to identify published evidence for development of AAAs in individuals with a non-aneurysmal scan. The following search strategy was employed for searching the PubMed database, resulting in the identification of 64 papers, of which three were potentially relevant:

abdominal aortic aneurysm [Title/ Abstract]  
AND normal [Title/ Abstract]  
AND scan OR screen\* [Title/ Abstract]

Sixty-one papers were excluded for a variety of reasons, described in Table 3.9. The three potentially relevant papers are detailed in Table 3.10.

<b>Reason for exclusion (other focus of paper)</b>	<b>No. papers</b>
Quality of life only	3

Imaging comparisons/ technique	7
Indicators for rupture	2
Other condition in AAA patients	8
Surgical issues	17
AAA not main focus	9
Biochemical aspects of AAA	4
AAA prevalence only	5
No AAA group not considered/ followed up	2
Review article only	2
Modelling article only	1
Earlier publication of relevant article	1
Relevant article	3
<i>Total</i>	<i>64</i>

Table 3.9 Excluded papers following systematic review for rupture in individuals with no AAA at scan

The differences in the results of these three studies are most likely due to differences in definitions and inclusions. Whilst the Huntingdon study follows up the largest number of individuals, new aneurysms are only counted where the second scan measurement is at least 0.5cm larger than the initial scan. Furthermore, the time between these two measurements would appear to vary between 6 months and around 7 years. This study also included a wide range of ages and consequently even with a mean 5.5 years between scans, the second scan may not have captured the majority of aneurysm development in the younger men. Both these factors suggest the estimation of aneurysm development in those with a normal initial scan is underestimated. The Gloucestershire study defined a normal scan as an aortic diameter of  $\leq 2.5$  cm, meaning that those with an initial aortic diameter in the range 2.6-3.0cm are not included in the calculation of aneurysm development in those with a normal initial scan; this is again likely to underestimate the number of growth transitions in this group. The results for growth in normals most applicable to the proposed model are those presented from a sub-set of the Chichester study. The estimates correspond to a group of 65-year-olds with a non-aneurysmal aorta defined as  $\geq 3.0$ cm, as employed in the proposed model. The rate of growth can be estimated from this study as 0.002 / 3-months, equivalent to a 3-month transition probability of 0.002 for the transition from the no AAA state to small AAA state in the model.

---

**Crow et al**

**Scott et al**

**Wilmink et al**

<b>Study based on</b>	Gloucestershire	Chichester	Huntingdon
<b>Age at initial scan</b>	65	65	45-95
<b>Definition of normal aorta</b>	≤ 2.5cm	≤ 3.0cm	≤ 3.0cm
<b>No. with no AAA at initial scan</b>	223	649	3606
<b>Timing of follow-up scans</b>	5, 12 years after initial scan	2,4,6,8,10 years after initial scan	Age ≥55 in June 1998 (commenced Nov 1991)
<b>No. person-years follow-up</b>	1863 (approx, assuming drop-out at last known scan)	3262 (approx, assuming drop-out at last known scan)	20013
<b>No. new aneurysms identified</b>	Unclear. Graph suggests five ≥3.0cm.	27 (≥3.0cm)	71 (expansion ≥0.5cm)
<b>Rate of growth transition</b>	2.7 / 1000py (approx)	8.3 / 1000py (approx)	3.5 / 1000py

Table 3.10 Details of papers identified as potentially relevant for estimating rupture in individuals with no AAA at initial scan

### 3.2.2 Rate of incidental detection

Since data was only collected for operations resulting from incidental detection of AAAs in MASS, the rate of incidental detection in those not in the recall screening cycle cannot be estimated directly from the trial. The evidence in the published literature is also sparse, as might be expected when considering the requirement for data collection about individuals *not* enrolled in a screening programme or trial.

As for growth in those with an initial non-aneurysmal scan, a limited systematic review of the literature was carried out to identify published evidence for incidental detection of AAAs outside screening programmes. The following search strategy was employed for searching the PubMed database, resulting in the identification of 16 papers, none of which were relevant:

abdominal aortic aneurysm [Title/ Abstract]  
 AND incidental OR opportunistic [Title/ Abstract]

All 16 papers were excluded for a variety of reasons, described in Table 3.11 below.

<b>Reason for exclusion (other focus of paper)</b>	<b>No. papers</b>
Individual case study only	6
Other condition in AAA patients	3
Potential for detection	1
Detection method if incidental finding of AAA	1
Prevalence of AAA if scanned for other reason	4
Letter only (no data presented)	1
Relevant article	0
<i>Total</i>	<i>16</i>

Table 3.11 Excluded papers following systematic review for incidental detection of AAAs

In light of the lack of available evidence regarding incidental detection, the data from controls in MASS was used indirectly to provide a crude estimate of the rate of detection in the four years of follow-up available. This calculation makes a number of assumptions about the number of elective operations via incidental detection in the control group and the known proportion of individuals with detected AAAs proceeding to elective surgery in the invited group to provide an estimate for the rate of incidental detection in the controls:

1.  $P(\text{elective operation} \mid \text{known AAA, invited attender}) = 305 / 1333$

2.  $P(\text{elective operation} \mid \text{known AAA, control}) = \frac{100}{x} = \frac{305}{1333}$

where  $x$  is the number of known AAAs in the control group (437), assuming the same proportion of known AAAs proceed to elective surgery in the incidentally detected group as in the screen-detected group.

3. Estimated AAAs (known and unknown) in invited group = 1333 screened + (6692 x 1333/27147) non-attenders + (329 x 1333/27147) non-visualised = 1678

Proportion of non-attenders and non-visualised attenders with AAA assumed to be the same as in visualised attenders. Expected total AAAs in control group assumed to be the same as total in invited group:  $y = 1678$

4.  $P(\text{detected} \mid \text{AAA, control}) = x / y = 437 / 1678$
5. 4-year  $P(\text{not detected} \mid \text{AAA, control}) = (1-p)^{16} = (1678-437) / 1678$

Where  $p$  is probability of an AAA being detected in a 3-month period, assuming the probability of detection remains constant over the four year period. Therefore the estimate of the three-month probability for incidental detection of AAAs for the model ( $p$ ) is 0.0187.

Furthermore, this is likely to provide an underestimation of the probability of detection based on the MASS figures, since it does not account for other removals from the “at risk” group via rupture or death (it is not feasible to account for these factors here, since the probability of these events depends on AAA size and other parameters included in the model). Consequently, this parameter estimate is only applied in a second version of the model (later referred to as Model A2), whilst the primary model (referred to as Model A1) instead considers incidental detection as the second parameter on which to calibrate (in addition to ruptures in large, undetected AAAs).

## **4 Validation against four-year MASS results**

The initial use of the model is to produce an estimate of four-year cost-effectiveness through the application of parameter estimates derived wherever possible directly from the MASS data. Estimation of the four-year results of the model can then be compared with the four-year results of the trial, allowing the origins of any model imprecision to be identified and giving an indication of overall performance. The total numbers of major events (such as operations and deaths) in each arm and the timing of these events over the four-year period are examined in addition to the estimates of life-years and costs. Discrepancies between the trial and model results are investigated and discussed, and a critical appraisal of the model and its derivation is given in conclusion.

### **4.1 Total events, benefits and costs**

In MASS, recruitment to the trial took place over three years, and a cut-off date (31<sup>st</sup> March 2002) was used to mark the end of follow-up for the four-year publication. Censoring due to staggered trial entry was therefore present over a three-year period, with the maximum follow-up for any individual of five years, and a mean of 4.1 years. In order to compare numbers of events from the trial with the model, censoring was applied to a five-year model, in the manner observed in the trial (see Section 2.4.2). The trial and model results in terms of numbers of events are given in Table 4.1. However, since the four-year cost results were presented for data truncated at four years and adjusted for censoring, the model results for total costs and life-years are taken from a four year model without censoring. Life-years and costs at four years are given in Table 4.2.

	<b>MASS</b>	<b>Model</b>
<b>Control</b>		
Elective operation	100	83
Emergency operation	62	62
Rupture	138	141
Contraindicated for elective surgery	nk	14
AAA death	113	109
Non-AAA death	3750	3724
<b>Invited</b>		
Elective operation		
Resulting from screen detection	295	282
Resulting from opp. detection	31	25
Emergency operation	28	34
Rupture	66	78
Contraindicated for elective surgery		
Resulting from screen detection	41	46
Resulting from opp. detection	nk	5
AAA death	65	69
Non-AAA death	3694	3724
Loss to recall follow-up, small/ med	304	289

nk = not known (relevant data not recorded in MASS)

Table 4.1 Numbers of key events in MASS and the model at four years

	<b>MASS</b>	<b>Model</b>
<b>Control</b>		
Costs (undiscounted)	£38.22	£36.05
Effects (undiscounted)	3.9933	3.9944
<b>Invited</b>		
Costs (undiscounted)	£103.67	£104.06
Effects (undiscounted)	3.9956	3.9961
<i>ICER* (annual discounting at 1.5% for effects, 6% for costs)</i>	<i>£28400</i>	<i>£37700</i>

\* incremental cost-effectiveness ratio (ICER); cost per life-year gained

Table 4.2 Cost and effect outcomes in MASS and the model at four years

Whilst the model performs well overall, there are some areas of discrepancy, notably ruptures (and consequently emergency operations and AAA deaths) in the invited arm. This is discussed in detail in Sections 4.3.1.

## 4.2 Events over time

In addition to comparing total numbers of events occurring during the four-year period, as described in Section 4.1, it is important to examine the performance of the model in terms of the timing of the events throughout the period. This will aid investigation and discussion of discrepancies as well as furthering the evidence for model performance. Results are presented for MASS, the censored model, and the uncensored model.

### 4.2.1 Elective operations

The parameter driving the estimation of elective operations in the unscreened population and the elective operations via incidental detection in the invited group is based on very little evidence (see Section 3.2.2). However, over four years, the model provides reasonable estimates of these events. There is a suggestion of some underestimation in the model near the start of the trial (see Fig 4.1); this is likely to be a result of the model assumption that no individuals already have an incidentally detected AAA at the outset. In practice, it is probable that a number of such individuals were not identified during pre-randomisation exclusions. Consequently there is a larger than expected number of operations early on, and the parameter estimate for the rate of incidental detection of AAAs over time is overestimated by the inclusion of these operations.

The model also provides good estimates for elective operations via the screening programme over time, see Fig 4.1. There is an apparently large difference between the model and the trial in the first cycle of the model, with a relatively large number of operations occurring in MASS and none in the model; in MASS, this represents many of the large aneurysms identified at initial screening undergoing elective intervention. Since the model assumes a one cycle (mean 45 day) delay for all individuals proceeding to elective surgery, this explains the operations occurring in the first cycle in MASS that are not represented in the model. However, the totals of the first two cycles combined are similar for the trial and the model (104 and 109 events respectively).

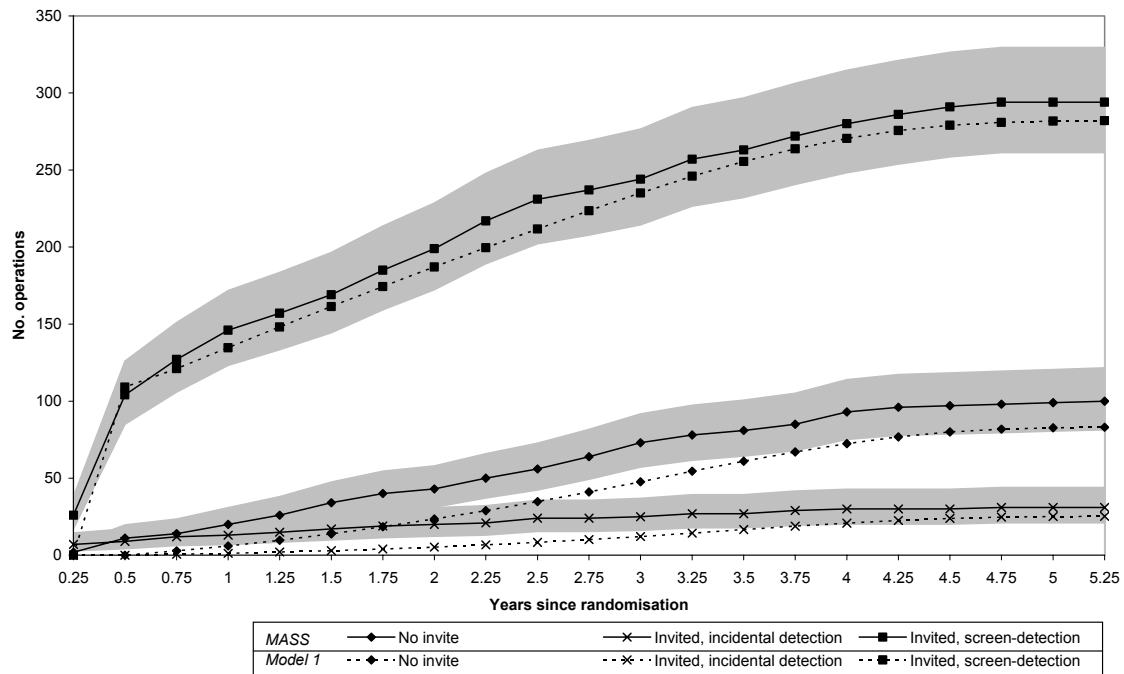


Fig 4.1 Cumulative elective operations over time for MASS and the model, with 95% confidence intervals for the observed data

## 4.2.2 Ruptures

The parameters for modelling ruptures are based on only small numbers of events in MASS, and chance variations in the number of these events in each three-month period in the trial results in apparent differences compared to the model. However, there are no systematic differences apart from a suggestion that a larger number of ruptures in the invited group occur early on in the trial than predicted by the model (see Fig 4.2). Although a constant rate of rupture is applied throughout the four year model, this rate may not be uniform over time, and it might be expected that the very large aneurysms present at the start of the trial are more likely to rupture than the generally smaller AAAs in the screened population later on (since the majority of large AAAs are removed from the population by elective intervention).

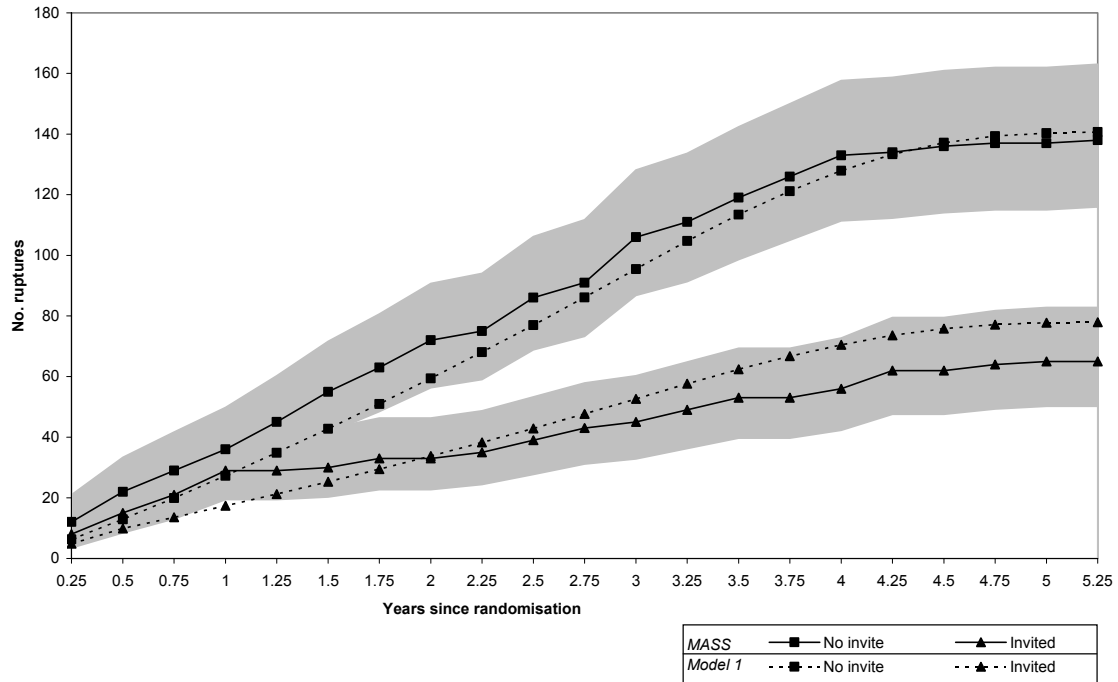


Fig 4.2 Cumulative ruptures over time for MASS and the model, with 95% confidence intervals for the observed data

### 4.2.3 Emergency operations

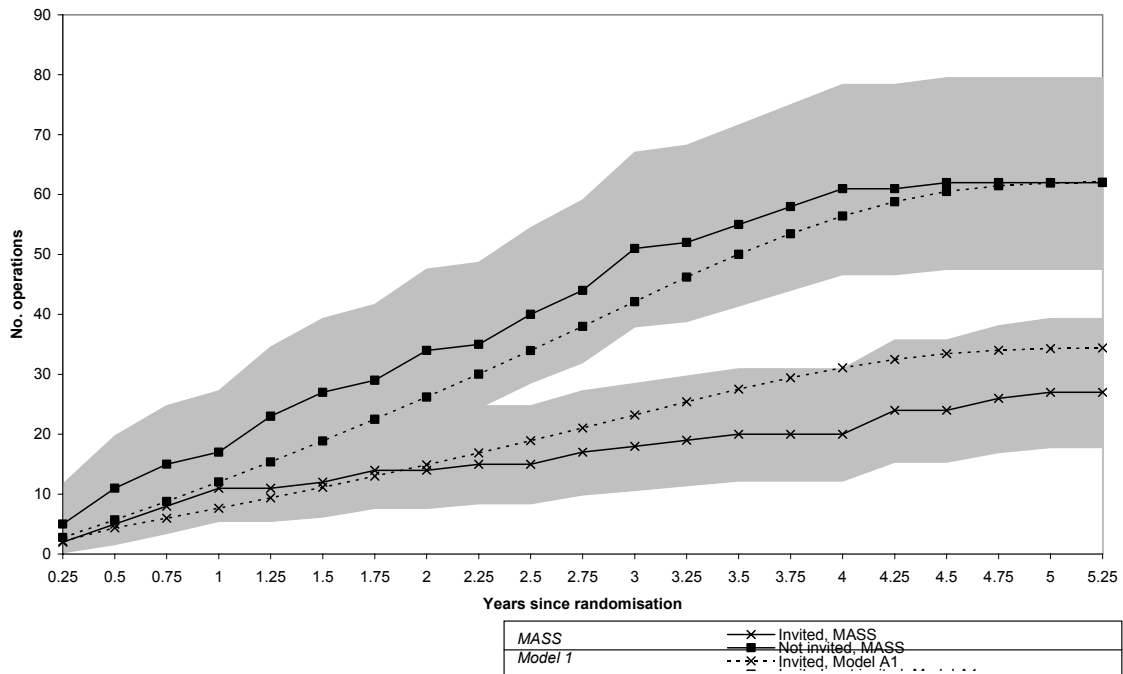


Fig 4.3 Cumulative emergency operations over time for MASS and the model, with 95% confidence intervals for the observed data

The occurrence of emergency operations over time (Fig 4.3) inevitably follows the pattern of ruptures over time. Consequently, a similar underestimation of events is seen in the early part of the model for the invited arm, but with no other systematic differences observed.

#### 4.2.4 AAA deaths

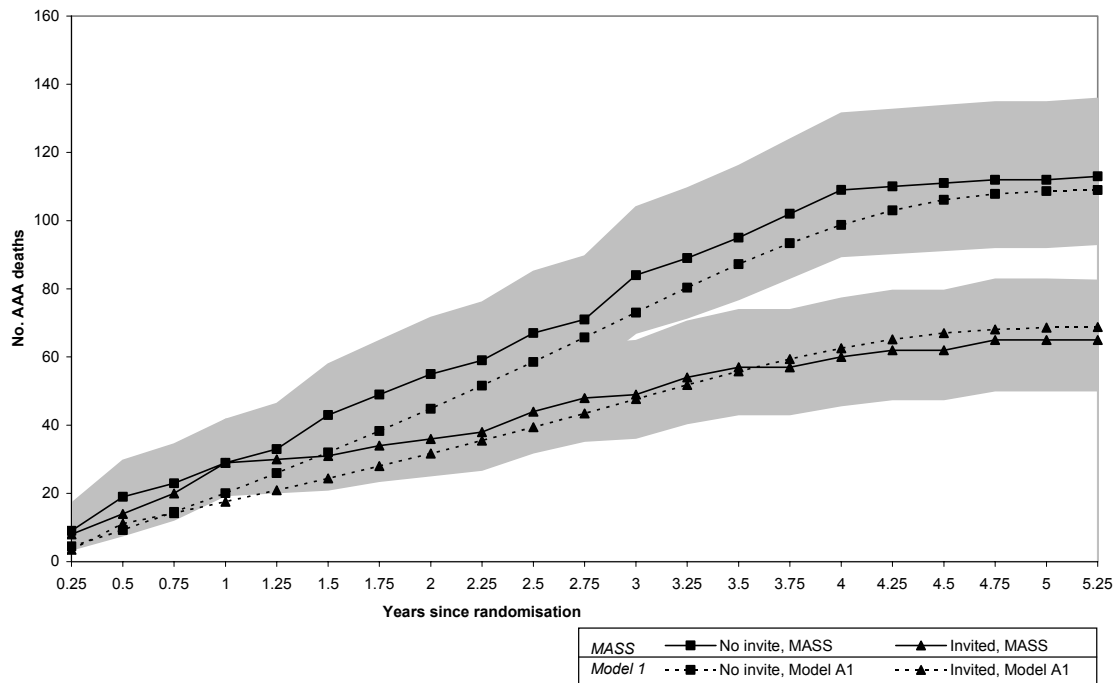


Fig 4.4 Cumulative AAA deaths over time for MASS and the model, with 95% confidence intervals for the observed data

There are again inevitable similarities with the pattern of AAA-related deaths over time (Fig 4.4) and ruptures over time. However, since the definition of an AAA-related death includes all deaths within 30 days of any AAA surgery, some of the excess deaths in the trial compared to the model can be explained by some elective operations occurring earlier in MASS (due to the one cycle delay to elective surgery imposed by the model structure, see Section 2.2.2). The relatively large number of AAA deaths occurring in the second cycle of the model is a result of all large AAAs detected at initial screening with a decision to proceed to elective surgery then receiving this operation in the second cycle. In MASS, these operations predominantly took place over the first two three-month periods (see Fig 4.1).

#### 4.2.5 Non-AAA deaths

In the four-year model, non-AAA death rates are taken directly from MASS (see Section 3.1.9). The probability of a non-AAA death is estimated using data from the control and invited groups combined, so the only small discrepancies between the trial and the model arise where the non-AAA deaths were unbalanced between the control and invited groups for a particular three-month period (Fig 4.5).

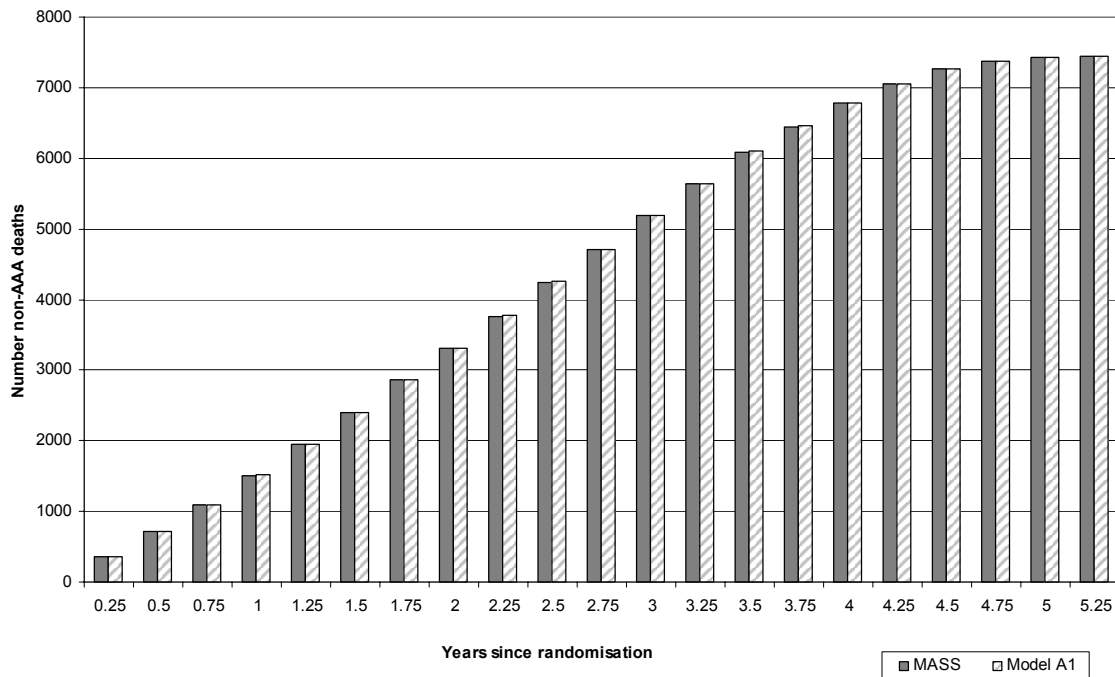


Fig 4.5 Cumulative non-AAA deaths in control and invited groups combined for MASS and the model

#### 4.2.6 Numbers in screen-detected AAA states

In addition to comparing major events in the trial and the model, it is also useful to examine the numbers in each of the screen-detected size states over time. It is not possible to make similar comparisons for the undetected or incidentally detected states, since the numbers in these states in the trial are not known. For the screen-detected states in the invited group however, the number of individuals in the small, medium and large states can be compared for three-month intervals following randomisation in the trial and at the end of each three-month cycle in the model. The data for the trial includes those returned to screening following a consultation for considering elective surgery in the

medium and large size states, since such individuals are included in these groups in the model. An individual in MASS with a measurement  $\geq 5.5$  cm at consultation but nevertheless returned to screening is assumed to be immediately returned to the large size state. Individuals with measurements  $< 5.5$  cm at consultation are returned to the medium size state.

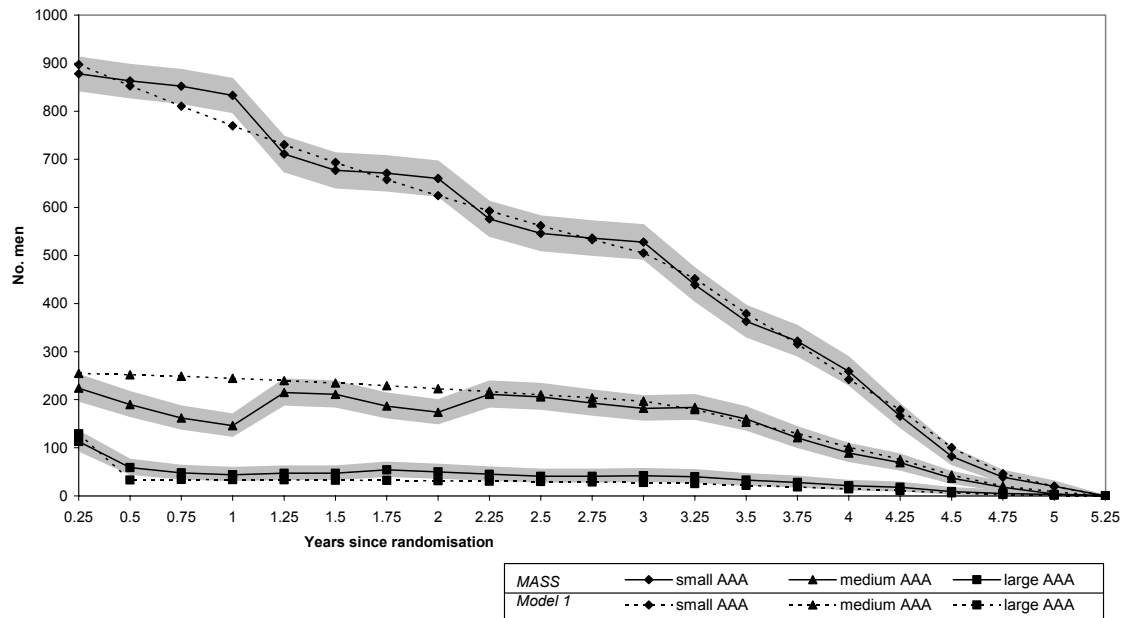


Fig 4.6 Number of individuals in screen-detected AAA size states in the invited group in MASS and the model

Numbers in the screen-detected small AAA size state are similar for the trial and the model over time (Fig 4.6). Small discrepancies can be seen over each one-year period since in MASS, recall scans for this group occurred annually, whereas the model represents this by a quarter of a scan in each three-month period.

For the medium detected state, the model represents the trial well later in the four year period, although there is considerable overestimation of numbers in this state early in the follow-up period (Fig 4.6). This is almost entirely due to the treatment of recall scans in the screen-detected small AAA group in the model; individuals growing to medium move to the medium size state in the same three-month period as the growth. In MASS, the identification as a medium-sized AAA did not occur until the annual recall scan. This explanation is confirmed by the similar numbers in the medium size state in period 5, representing the first annual scan for small AAAs in MASS.

Numbers in the large detected state are reasonably well represented by the model (Fig 4.6). The relatively large number in this state in the first three-month period in both the trial and the model reflects the delay to elective surgery for patients identified for this procedure.

## **4.3 Conclusions and discussion**

### **4.3.1 Model validation**

Over the four-year comparison period, the model appears to perform well in general, both in terms of overall costs and life-years, and numbers of events over time. A small number of areas giving rise to discrepancies have been identified however, and it is important to consider the significance of these, particularly for the long-term extrapolation of this proposed model. This includes numbers in the medium detected state, and the number and timing of ruptures in the invited arm; these will now be discussed in turn.

The timing of entry into the medium detected state accounts for most of the differences in numbers in this state in the trial and the model (see Fig 4.13). The majority of the additional person-years in this state in the model are accrued early in the follow-up period; over time, there are less individuals in all detected states as intervention and ruptures take place. This reduces the discrepancy over time, and hence will not greatly influence the long-term model.

The overestimation of ruptures (and resulting overestimation of emergency operations and AAA deaths) in the invited arm is perhaps the most serious difference between the model and trial. It results from overestimation of ruptures in one sub-group of the invited group: non-attenders (either at the initial scan, or those lost to recall follow-up; see Table 4.3). The parameter determining ruptures in this group (*rr\_grp4un*) is also used to calculate ruptures in the majority of (undetected) large AAAs in the control arm of the model. Estimation of this parameter was problematic, and the value applied in the model was only chosen because it resulted in a close match of the MASS data in the control arm

results. An alternative structure would be to use two separate parameters for rupture in large undetected AAAs for those in the invited arm (i.e. non-attenders) and for those in the control arm (a mixture of the individuals accounted for in the large detected, large undetected and contraindicated groups in the invited arm).

	<b>Ruptures, invited arm</b>	
	<b>MASS</b>	<b>Model</b>
Non-aneurysmal	6	7
Small detected AAA in clinical follow-up	1	1
Medium detected AAA in clinical follow-up	7	8
Large detected AAA in clinical follow-up	14	12
Contraindicated for elective surgery	9	11
Non-attenders	27	39
Non-visualised*	2	-
<b>Total</b>	<b>66</b>	<b>78</b>

\* Distributed across all other categories in model, see Section 2.2.1

Table 4.3 AAA sizes at which ruptures occur in MASS and the model

In addition to the misrepresentation of ruptures in large undetected AAAs in the model, there is also a difference in timing of ruptures across all states in the invited arm. The use of a constant probability of rupture in each state over time may not be appropriate (see Fig 4.5). Parameters for some subgroups (such as large detected AAAs) where estimation is influenced by a relatively large number of events in the first year of the trial may over-estimate events in the long term. Other parameters may also be subject to change over the course of the screening programme; the probability of incidental detection or drop-out may not remain constant (as discussed in Section 3.1.4). Furthermore, the affect of age on the majority of parameters in the model may be significant in indicating the need for time-dependent parameters in order to obtain valid inferences from a life-time model for cost-effectiveness.

The costs in the model are similar to the trial although it is known that additional costs are included in the invited arm of the model as a result of costing fractions of recall scans (see Section 2.4.3). However, these additional costs are in part compensated for by the absence of costs arising from consultations following referral for AAA expansion (in MASS, this accounted for 152 consultations during the mean four-year follow-up period).

In conclusion, the principal limitations for the long-term modelling relate to the estimation of the rupture parameters and the consequent numbers and timing of ruptures, and the potential impact of time and age on parameters currently applied as time-independent.

#### **4.3.2 Model development**

The data-driven aspects of model development (calibration based on the parameter values for ruptures in large, undetected AAAs and incidental detection) also have implications for the validity of the model. It is inevitable that the MASS trial results can be closely replicated by a model employing parameters derived from the same dataset, particularly in areas where parameters are chosen to provide a good fit to the data (ruptures in large, undetected AAAs and incidental detection). However, this may not mean that the model is an accurate representation of the underlying process, and this will become apparent when extrapolated. The model will therefore be further tested by validation on three further years of follow-up in the MASS trial and on external data on AAA screening.

