# The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm

**Helen Campbell, Andrew Briggs, Martin Buxton[1], Lois Kim[2], Simon Thompson[2]**

Health Economics Research Centre, Department of Public Health, University of Oxford; [1]Health Economics Research Group, Brunel University, Middlesex; [2]MRC Biostatistics Unit, Institute of Public Health, University of Cambridge, Cambridge, UK

***Objectives*: To review health economic models of population screening for abdominal aortic aneurysm (AAA) among elderly males and assess their credibility for informing decision-making.**

***Methods*: A literature review identified health economic models of ultrasound screening for AAA. For each model focussing on population screening in elderly males, model structure and input parameter values were critically appraised using published good practice guidelines for decision analytic models.**

***Results*: Twelve models published between 1989 and 2003 were identified. Converting costs to a common currency and base year, substantial variability in cost-effectiveness results were revealed. Appraisals carried out for the nine models focusing on population screening showed differences in their complexity, with the simpler models generating results most favourable to screening. Eight of the nine models incorporated two or more simplifying structural assumptions favouring screening; uncertainty surrounding these assumptions was not investigated by any model. Quality assessments on a small number of parameters revealed input values varied between models, methods used to identify and incorporate input data were often not described, and few sensitivity analyses were reported.**

***Conclusions*: Large variation exists in the cost-effectiveness results generated by AAA screening models. The substantial number of factors potentially contributing to such disparities means that reconciliation of model results is impossible. In addition, poor reporting of methods makes it difficult to identify the most plausible and thus most useful model of those developed.**

## Introduction

In the UK, decision analytic models are increasingly used to estimate the costs and outcomes of alternative health care interventions. Cost-effectiveness data presented to the National Institute for Health and Clinical Excellence (NICE), the body responsible for appraising health technologies on behalf of the UK National Health Service (NHS), are largely model based.[1]

The appeal of decision modelling is its versatility. Its uses include extrapolation beyond trial outcomes, evidence synthesis, generalization of study results to alternative settings and identification of future research priorities.[2] However, concerns about decision models (notably a lack of methodological rigour and transparency) have been raised and policy-makers still regard decision model findings with caution.

Population ultrasound screening for abdominal aortic aneurysm (AAA) in elderly males provides a good example. The effectiveness and cost-effectiveness of a national screening programme for AAA has been debated in many countries over many years.[3,4] First attempts around 1990 to determine the likely cost-effectiveness of a screening policy involved some simple modelling studies.[5,6] By 2000, the number of models had increased substantially. A large pragmatic randomized trial of AAA screening in the UK provided the first trial-based estimates of effectiveness and resource-use in 2002.[7] But with follow-up limited to four years, analysts have continued to develop models to estimate long-term cost-effectiveness. However, there is little evidence that these models have assisted decision-makers in determining whether or not screening for AAA is cost-effective. As of October 2005, longstanding decisions to refrain from screening all elderly males,

**Helen Campbell MSc,** Research Officer, **Andrew Briggs DPhil,** Reader in Health Economics, Health Economics Research Centre, Department of Public Health, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF; **Martin Buxton BA,** Professor of Health Economics, Health Economics Research Group, Brunel University, Uxbridge, Middlesex; **Lois Kim MSc,** Research Associate, **Simon Thompson DSc,** Professor of Biostatistics, MRC Biostatistics Unit, Institute of Public Health, University of Cambridge, Cambridge, UK.

Correspondence to: helen.campbell@dphpc.ox.ac.uk

which were made in the absence of definitive cost-effectiveness evidence, continue to be upheld in the United States,[3] Canada,[4] and the UK.

That policy-makers appear unable or unwilling to make recommendations for (or against) AAA population screening on the basis of results generated predominantly by decision analytic models, raises a number of important questions. The aim of this paper was to identify and review published AAA screening models, with a view of determining their credibility for informing policy.

## AAAs and screening

AAAs are present in 5–7% of men over 65 and account for approximately 2% of deaths in this group in the UK. In most cases, the aortic dilation is asymptomatic, going undetected until rupture. Prognosis following rupture is bleak as around half of patients will die before they reach hospital and the remaining half will face emergency surgery mortality rates ranging from 30% to 70%. Ultrasound is a low-cost, non-invasive, sensitive and specific screening test for early detection of AAA. Following detection, AAAs can be monitored for growth, and patients referred for elective surgical repair when the aneurysm exceeds a size beyond which the probability of rupture is considered high.

## Methods

Full details of the terms used and databases included in the search for health economic models of AAA screening are available as a web appendix together with other methodological details. The web appendix may be viewed free of charge at http://www.ingentaconnect.com/content/rsm/jhsrp/2007/00000012/00000001/art00005. All fields within databases were searched in February 2005 for papers or reports published up to and including 2004, with no language restrictions. Papers reporting health economic models were reviewed, and costs expressed as 2003/2004 £ Sterling to facilitate a comparison of results.

Only models for population screening among elderly males were critically appraised. The structure of each model was assessed using guidelines for good practice in decision analytic modelling in health technology assessment (Table 1).[8] An appraisal of the data used to inform three key parameters (a clinical parameter, a cost parameter and an outcome parameter) common to each model was also carried out. For each, an assessment was made of the methods used to (1) identify, (2) incorporate and (3) handle uncertainty around model input data.[9] Where papers had referenced another publication as the source of input data, this publication was retrieved and reviewed.

## Results

Twelve health economic models of AAA screening were identified. Table 2 presents base-case results reported or derivable for each model. Where studies reported results for more than one screening strategy, or for different patient groups, those pertaining to a single ultrasound screen, and to all elderly males, were sought. Where estimates of the additional cost per patient resulting from AAA screening were reported, values ranged from £65 to £460. Effect differences ranged from –0.001 to 0.28 among studies reporting life years and 0.031 to 0.077 among those reporting quality–adjusted life years (QALYs). Incremental cost per life year gained figures ranged from −£101,443 to £35,187, and incremental cost per QALY figures from £817 to £7,738.

A detailed assessment of the structure and input parameter values of nine of the models in Table 2 was undertaken. The remaining three models were excluded (grey shading in Table 2), two because their focus was targeted ultrasound screening for familial AAA,[10,11] and one because analysts failed to report any of the statistics in Table 2.[12]

**Table 1** Issues for consideration with regard to model structure[8]

| No. | Heading | Description |
|---|---|---|
| 1 | The decision problem/ objective | • The problem needing to be addressed, i.e. the disease/condition involved, the patient group, and the intervention to be evaluated<br>• The objective of the evaluation/model<br>• The primary decision-maker. |
| 2 | The study perspective/scope | • The perspective adopted for the analysis. Model inputs should be consistent with this perspective |
| 3 | Rationale for model structure | • The model structure should represent the underlying disease process, and the impact of the intervention<br>• Sources of data used to develop the model structure should be reported |
| 4 | Structural assumptions | • Assumptions made while structuring the model – these should be justified and be reasonable<br>• Uncertainty associated with structural assumptions should be examined using sensitivity analysis |
| 5 | Strategies/comparators | • Alternative strategies being evaluated |
| 6 | Model type | • The model type used, e.g. decision tree, Markov model. This will be determined by the decision problem |
| 7 | Time horizon | • The length of follow-up in the evaluation – this should extend far enough into the future to capture all costs and consequences associated with the interventions being compared |
| 8 | Disease states/pathways | • Disease states or pathways through the model – these should reflect important events in the underlying process of the disease/condition |
| 9 | Cycle length | • In discrete time models – the duration of time before the model cycles. This period should correspond to the minimum period over which a change in disease status could be expected to occur |

**Table 2** Baseline costs, effects and cost-effectiveness results from 12 models (ordered by date of publication) comparing ultrasound screening for AAA versus no screening (all costs are expressed in 2003/04 £UK)

| | **Bengtsson *et al.*** [*,†] | **Russell** [†,‡] | **Collin** | **Mason *et al.*** | **Frame *et al.*** [§] | **Law *et al.*** [†] | **St Leger *et al.*** | **Lee *et al.*** | **Boll *et al.*** | **Pentikainen *et al.*** [¶,**] | **Soisalon-Soininen *et al.*** [**] | **Connelly *et al.*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study reference number | 19 | 5 | 18 | 14 | 15 | 6 | 13 | 16 | 17 | 10 | 11 | 12 |
| Publication year | 1989 | 1990[††] | 1990[††] | 1993 | 1993[††] | 1994 | 1996 | 2002[††] | 2003 | 2000[††] | 2001 | 2002 |
| Country | Sweden | UK | UK | UK | US | UK | UK | US | Netherlands | Finland | Finland | Canada |
| *Costs* | | | | | | | | | | | | |
| Mean cost per patient with screening | £151 | NR | £241 | NR | £783 | NR | £124 | £1202 | £412 | NR | £749 | NR |
| Mean cost per patient without screening | £55 | NR | £176 | NR | £581 | NR | £15 | £742 | £153 | NR | £363 | NR |
| *Outcomes* | | | | | | | | | | | | |
| Mean life years per patient with screening | 0.00748[‡‡] | NR | NR | NR | 8.5704 | NR | NR | NR | 17.27 | NR | NR | NR |
| Mean life years per patient without screening | 0.02549[‡‡] | NR | NR | NR | 8.5647 | NR | NR | NR | 16.99 | NR | NR | NR |
| Mean QALYs per patient with screening | NR | NR | NR | NR | NR | NR | NR | 9.944 | NR | NR | NR | NR |
| Mean QALYs per patient without screening | NR | NR | NR | NR | NR | NR | NR | 9.884 | NR | NR | NR | NR |
| *Cost-effectiveness* | | | | | | | | | | | | |
| Mean cost difference per patient | £96 | NR | £65 | £101 | £202 | £115 | £109 | £460 | £260 | £404 | £385 | NR |
| Mean life year difference per patient | 0.01801 | NR | NR | –0.00100 | 0.00570 | 0.08354 | NR | NR | 0.28 | 0.07000 | 0.09200 | NR |
| Mean QALY difference per patient | NR | NR | 0.07692 | NR | NR | NR | NR | 0.03132 | 0.059 | NR | NR | NR |
| Incremental cost per life year gained | £5332 | NR | NR | –£101,433 | £35,511 | £1382 | NR | NR | £1573 | £5776 | £4187 | NR |
| Incremental cost per QALY gained | NR | £817 | £842 | NR | NR | NR | £3494 | £7738 | NR | NR | NR | NR |

Studies shaded in grey were not critically appraised for reasons outlined in the text. NR, not reported or easily derivable
Incremental cost per life year gained discounted, mean costs and effects not discounted
*Results presented separately for all elderly men, and for men with intermittent claudication. Those shown here are for all elderly males
[†]Policy of repeat screening modelled
[‡]Results presented separately for men and women operated on at different ages. Those shown here relate to men aged 60
[§]Results presented separately for a single ultrasound screen, an initial ultrasound screen plus repeat screen at five years, and screening using abdominal palpation. Those shown here are for a single ultrasound screen
[¶]Results presented separately for men and women. Those shown here relate to men
**Not population screening models
[††]No financial year reported for costs – assumed to be the year prior to publication
[‡‡]Absolute effect measured as life years lost with and without screening
[§§]Incremental cost per life year gained discounted, mean costs and effects not discounted

## Model structures

All papers had a stated objective to estimate the cost-effectiveness of a population ultrasound screening programme for AAA compared with no screening. Only two papers provided details of funding sources, and none indicated whether the modelling had been commissioned directly by a health care provider or conducted independently. The study viewpoint was explicitly stated in only one paper,[13] although all models used inputs consistent with a health care provider's perspective.

Consideration of the strategies being modelled showed the age at which men were initially screened varied from 60 to 79. In six papers,[13–18] a policy of a single (or prevalence) population screen was modelled and in three papers, policies involving multiple population screens were presented.[5,6,19] If estimates of the annual incidence of new AAA of 0.1% per annum are to be believed,[15,19] then the prevalence of AAA at subsequent screens will be small. In terms of cost-effectiveness, and when comparing against a policy of no screening, one might expect a policy involving re-screens of the same cohort to appear less favourable than a single screen strategy. Table 2, however, shows that incremental cost-effectiveness ratios (ICERs) from two of the three studies were among the three most favourable reported.[5,6]

Cross-referencing model type with the results in Table 2 revealed that the three studies reporting the lowest ICERs[5,6,18] were also the three not employing a decision analytic framework to model explicitly costs and effects. Although a number of terms (including mathematical model, computer spreadsheet model and Markov model) were used to describe the remaining six models, each used Markov modelling. The underlying disease process appears to provide the rationale for the structure of these six models, with the health states included in each model reflecting important clinical stages in the disease process (e.g. no AAA, small, medium and large AAA). Each model appears to cycle on a yearly basis; however, in only one paper,[17] this is stated explicitly.

In terms of the time horizon modelled, five studies conducted a lifetime analysis,[5,14,16–18] three used 20 years,[6,15,19] and one used 15 years.[13] Given the starting age of the patient cohort, 15 or 20 years might be considered broadly equivalent to a lifetime analysis and therefore time horizon is unlikely to be a significant factor contributing to disparate results.

Eight of the nine models incorporated two or more simplifying structural assumptions favouring screening, and so cost-effectiveness results could be overly optimistic. Four of the nine models appear to assume that without screening, opportunistic detection of AAA and resultant elective repairs would not occur.[5,6,13,14] Although data on AAA diagnosis in the absence of screening are scarce, patients still undergo elective AAA repair in the absence of formal screening programmes. Omitting some level of natural case finding will over-estimate to some degree lives saved and life years gained by implementation of such screening. In a one-way sensitivity analysis conducted by one of the five models, which allowed for opportunistic detection, reduction of the base-case estimate from 6.6% per year to zero had only a small effect on the ICER.[16]

Four of the nine models seem to be structured for 100% attendance at screening.[6,15,16,19] Pilot studies of AAA screening have suggested attendance of around 80%. An assumption that all invited patients attend will tend to make screening appear more favourable than it would be in routine practice. Ideally models should reflect the positive association between attendance rates and invitation costs (i.e. only re-invitation and persistent follow-up of non-responders can improve turnout). Table 2 shows that despite assuming 100% take-up of screening, three of these four studies still generated the highest ICERs. One-way sensitivity analysis by one of the models structured for non-attendance suggested only small reductions in cost-effectiveness as attendance declines.[17]

All nine analyses assume ultrasound sensitivity and specificity to be 100%. While the test is unlikely to be perfectly accurate, evidence suggests its sensitivity and specificity are close to 100%.[20] Given the uniformity of this assumption, it cannot be a factor contributing to the between-model differences in cost-effectiveness results.

Assumptions about the impact of screening upon AAA rupture vary between studies. Table 3 shows that one study assumes that screening will completely eliminate rupture and emergency surgery.[19] In addition to 100% attendance at screening and 100% sensitivity and specificity of ultrasound, two further assumptions would need to hold in order for this to be true: firstly no ruptures in patients with screen-detected AAA below the threshold for surgery, and secondly all patients with screen-detected AAA exceeding this threshold undergo elective repair without delay. As the threshold for surgery used in this study is small (40 mm) and rupture of an AAA of this size is unlikely, the first of these two assumptions could be considered valid. That all patients indicated for elective repair will undergo the procedure is, however, improbable, although Table 3 shows a further two studies also making this assumption.[6,16] Only five of the papers acknowledged that not all patients exceeding the threshold for elective repair would undergo surgery, either because they refused or were contra-indicated for the procedure, or because they failed to attend the initial screen and their AAA remained undetected.[5,13–15,17]

Table 3 suggests that the size of aneurysm threshold beyond which elective repair is indicated has increased over time. The four most recent studies, which use larger diameter thresholds, all acknowledge a small risk of rupture below this size. It is difficult to determine the impact of these assumptions upon cost-effectiveness results. From Table 3, it could be inferred that two models[13,17] simulate the most likely impact of screening

**Table 3** Threshold sizes for elective surgery and structural assumptions relating to rupture and emergency surgery in the screening arms of models, ordered by date of publication

| Study | <Threshold | | >Threshold |
|---|---|---|---|
| | Model structured to allow rupture and emergency surgery in screening arm | Threshold for surgery | Model structured to allow rupture and emergency surgery in screening arm |
| Bengtsson *et al.* (1989)[19] | No | 40 mm | No |
| Russell (1990)[5] | NA | 40 mm | Yes* |
| Collin (1990)[18] | NA† | 40 mm‡ | NA† |
| Mason *et al.* (1993)[14] | No | 50 mm | Yes |
| Frame *et al.* (1993)[15] | No | 40 mm | Yes |
| Law *et al.* (1994)[6] | Yes§ | 60 mm | No |
| St Leger *et al.* (1996)[13] | Yes | 60 mm | Yes |
| Lee *et al.* (2002)[16] | Yes | 50 mm | No |
| Boll *et al.* (2003)[17] | Yes | 50 mm | Yes |

*Text accompanying model acknowledges that not all patients identified as surgery candidates will undergo elective AAA repair. Whether ruptures occurring in this group are fed through into the modelling is unclear
†It is acknowledged that rupture and emergency surgery will still occur with screening, but the AAA size at which rupture is permitted is not specified
‡40 mm mentioned in text as a likely threshold for surgery, inclusion in the model is unclear
§Text accompanying model acknowledges that rupture will still occur below the threshold for surgery. Whether this is fed through into the modelling is unclear
NA – information not available

upon AAA rupture, since only they permit rupture both below and above the threshold. Table 2 shows, however, that the results from these models were among the most favourable towards screening.

## Elective AAA repair mortality rate

Base-case estimates of this parameter varied between studies. (Table 2 on web) The two studies using the lowest rates[5,18] produced the lowest cost-effectiveness ratios. Higher mortality rates, however, were not always associated with less favourable results. Only two studies provided details of bibliographic searches to identify papers to inform this parameter value.[15,17] Three studies provided no reference at all for the base-case estimate used.[5,16,18] Of the six studies that did cite sources, three referenced just one publication[13–15] (although one of these contained results of a Medline review), and one each referenced two papers,[17] three papers,[6] and four papers.[19] Of those studies citing more than one source, none provided information on how estimates from these studies were synthesized to generate the base-case value incorporated within the model.

Four of the nine studies reported sensitivity analyses for this parameter. Of the three performing one-way analyses, two demonstrated that small changes in mortality following elective AAA repair would have a large impact upon cost-effectiveness.[14,19] The reporting of sensitivity analysis results from the third study, however, did not use the ICER.[13] The remaining paper carrying out sensitivity analysis for this parameter did so as part of a multivariate analysis.[15]

## Cost of emergency AAA repair

The assumed cost of emergency AAA repair varied substantially across the nine models. (Table 3 on web) Around half of the analysts used a base-case cost that they had estimated themselves[16–19] and in one paper both a locally calculated cost and a published cost were used.[13] For just two of these studies, detail was available (in a further published paper) on resource-use included in the estimate, and costing methods used.[13,16] Of the remaining four studies, cost estimates used by three appear to be based on expert clinical opinion[5,6] or personal communication[14] and one study referenced a single paper reporting a costing exercise.[15] Details on the costing methods employed and resource-use included in this estimate were not, however, reported.

Four of the nine studies reported sensitivity analyses for the cost estimate of emergency AAA repair. Two of these studies undertook one-way sensitivity analyses;[13,19] however, it was possible to determine the impact of variation in the cost estimate on the ICER for only Bengtsson *et al.*[19] who found that doubling the emergency surgery cost reduced their base-case incremental cost per life year by 30%. Of the remaining two studies, one reduced the costs of elective and emergency repair to the same value simultaneously,[14] while the impact of varying the cost of emergency AAA repair in the remaining paper was again assessed at the same time as other parameter values were varied.[15]

## Utility levels assigned to life years modelled

Of the four studies using the QALY as their outcome measure, two assigned a utility level of one to all life years modelled.[13,18] (Table 4 on web) Neither of these studies provide justification for such an assumption, nor do they examine uncertainty surrounding this value. Perfect health was also generally assumed for patients in another model; however, utility for AAA repair survivors in this model was reduced slightly in the three-month period following surgery to reflect

procedure-related morbidity.[5] Disutility associated with AAA repair was also incorporated within the model by Lee *et al.*,[16] as were utility levels associated with possible long-term complications. Again no sources were provided for utility values used and no sensitivity analyses exploring the impact of alternative values were conducted.

## Discussion

There is a lack of agreement between models, which raises questions about the overall quality of the modelling employed. Any decision-maker attempting to review these models would be confronted with poor reporting of results. In only four papers, it was possible to derive estimates of the mean per patient costs and outcomes with and without screening. No improvement in the standard of reporting over time was apparent.

Although 10 models generated ICERs well below £20,000 (which is likely to be considered cost-effective in most jurisdictions),[21] considerable uncertainty surrounds the additional costs that would need to be incurred to generate additional health benefit. Given the substantial number of factors with the potential to influence model results, attributing these differences to specific modelling techniques, structural assumptions or parameter values is simply not possible. We can only speculate about the possible relative impact of different study components. Model type, for example, could be a contributing factor – none of the three studies generating results most favourable to screening constructed a model framework to model costs and effects explicitly.

Examination of structural or simplifying assumptions revealed eight of the nine population screening models to have incorporated at least two assumptions, which would artificially favour a screening programme. Although consideration of each assumption in isolation suggested the likely impact in terms of overestimating cost-effectiveness might be small, the collective impact of these assumptions on the results of each model is uncertain, and effectively impossible for a decision-maker to determine. It might be concluded that Law *et al.*[6] incorporated the most number of assumptions in favour of screening and that this resulted in a model generating one of the lowest ICER values. In contrast, the model by Boll *et al.*[17] incorporated the least number of structural assumptions in favour of screening. However, it too generated results highly favourable towards screening.

One other possible source contributing to the divergent results are the data used to populate these models. Having to compare and ascertain the quality of all input parameter values across all nine models highlights the enormity of the task facing the decision-maker. Assessing data quality for just three model parameters revealed a wide range of input values for each. The relation with reported results was not always intuitive, given the base-case parameter values reported.

Disparate results among models evaluating the same interventions have been observed elsewhere.[22–24] When convergent validity between models is low, and reconciliation of disparate results not achievable, the decision-maker will need to determine whether results from any of the models are robust enough to inform policy. For AAA screening models, determining the reliability of model input parameters and consequently of cost-effectiveness results is simply not possible. Good practice guidelines for decision analytic models suggest that the analyst should document all the information sources that have been searched.[9] Such details are rarely reported. Without this information, the decision-maker cannot judge whether parameter values are appropriate. A lack of methodological rigour in reporting search strategies for identifying data to populate decision models has also recently been observed by Cooper *et al.*[2] in decision models developed for UK Health Technology Assessments. They found that for model parameters (with the exception of clinical effectiveness data), methods used to identify sources of evidence were rarely reported and appeared to be *ad hoc*.

In the absence of suitably robust decision models, interaction between researchers and decision-makers could provide an opportunity to improve both the transparency and usefulness of published analyses. In the UK, for example, NICE commissions work and interacts with analysts, which facilitates the modelling of alternative scenarios.

Despite not being able to determine whether models were commissioned directly by decision-makers or conducted by independent analysts, we assumed that models were devised so as to meet the requirements of a decision-maker and would have been known about and considered at the time of policy review. Input parameters were certainly consistent with a health service viewpoint, suggesting the first of these assumptions to be plausible. Furthermore, a recently published synthesis showed that decision-makers systematically search for and identify the majority of published models.[3]

## Conclusion

There are a number of reasons why cost-effectiveness models may not have provided an adequate basis to encourage policy-makers to adopt population screening for AAA. First, convergent validity between models is low. Second, it is extremely difficult based upon the data reported in published studies to attribute differences in results to one or more particular sources. Third, poor reporting of methodology makes it difficult to ascertain whether the modelling carried out is of a sufficiently high standard to inform policy-making. There need to be major improvements in the construction and reporting of health economic decision

models if they are to contribute to health policy decisions.

## Acknowledgements

The study was funded by the UK NHS National Screening Committee.

## References

1 Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet* 2002;**360**:711–5
2 Cooper NJ, Coyle D, Abrams KR, Mugford M, Sutton AJ. Use of evidence in decision models: an appraisal of Health Technology Assessments in the UK since 1997. *J Health Serv Res Policy* 2005;**10**:245–50
3 Meenan RT, Fleming C, Whitlock EP, Beil TL, Smith P. Cost-effectiveness analyses of population-based screening for abdominal aortic aneurysm: evidence synthesis. In: *Report of the US Preventive Services Task Force. Guide to Clinical Preventive Services*. 3rd edn. Periodic Updates. Rockville, MD: Agency for Healthcare Research and Quality, 2005
4 Patterson C. *Screening for Abdominal Aortic Aneurysm*. Ontario, Canada: Canadian Task Force on Periodic Health Examination, 1994
5 Russell JGB. Is screening for abdominal aortic aneurysm worthwhile? *Clin Radiol* 1990;**41**:182–4
6 Law MR, Morris J, Wald NJ. Screening for abdominal aortic aneurysm. *J Med Screen* 1994;**1**:110–6
7 Multicentre Aneurysm Screening Study Group. Multicentre aneurysm screening study (MASS): cost-effectiveness analysis of screening for abdominal aortic aneurysms based on four year results from randomised controlled trial. *BMJ* 2002;**325**:1135–41
8 Phillips Z, Ginnelly L, Sculpher M, *et al.* Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**:36
9 Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics* 2000;**17**:36
10 Pentikainen TJ, Sipila T, Rissanen P, Soisalon-Soininen S, Salo J. Cost-effectiveness of targeted screening for abdominal aortic aneurysm. Monte Carlo-based estimates. *Int J Technol Assess Health Care* 2000;**16**:22–34
11 Soisalon-Soininen S, Rissanen P, Pentikainen T, Matilla T, Salo JA. Cost-effectiveness of screening for familial abdominal aortic aneurysms. *Vasa* 2001;**30**:262–70
12 Connelly JB, Hill GB, Millar WJ. The detection and management of abdominal aortic aneurysm: a cost-effectiveness analysis. *Clin Invest Med* 2002;**25**:127–33
13 St Leger AS, Spencely M, McCollum CN, Mossa M. Screening for abdominal aortic aneurysm: a computer assisted cost utility analysis. *Eur J Vasc Endovasc Surg* 1996;**11**:183–90
14 Mason JM, Wakeman AP, Drummond MF, Crump BJ. Population screening for abdominal aortic aneurysm: do the benefits outweigh the costs? *J Publ Health Med* 1993;**15**:154–60
15 Frame PS, Fryback DG, Patterson C. Screening for abdominal aortic aneurysm in men ages 60 to 80 years. A cost-effectiveness analysis. *Ann Intern Med* 1993;**119**:411–6
16 Lee TY, Korn P, Heller JA, *et al.* The cost-effectiveness of a "quick screen" program for abdominal aortic aneurysm. *Surgery* 2002;**132**:399–407
17 Boll APM, Severens JL, Verbeek ALM, van der Vliet JA. MASS screening on abdominal aortic aneurysm in men aged 60 to 65 years in the netherlands. impact on life expectancy and cost-effectiveness using a Markov model. *Eur J Vasc Endovasc Surg* 2003;**26**:74–80
18 Collin J. The value of screening for abdominal aortic aneurysm by ultrasound. In: Greenholgh RM, Mamaid JA, eds. *The Causes and Management of Aneurysms*. Philadelphia, PA: Saunders, 1990
19 Bengtsson H, Bergqvist D, Jendteg S, Lindgren B, Persson U. Ultrasonographic screening for abdominal aortic aneurysm: analysis of surgical decisions for cost-effectiveness. *World J Surg* 1989;**13**:266–71
20 Lindholt JS, Vammen S, Juul S, Henneburg EW, Fasting H. The validity of ultrasonographic scanning as screening methods for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg* 1999;**17**:472–5
21 National Institute for Clinical Excellence. *Guide to The Methods of Technology Appraisal*. London: NICE, 2004
22 Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde D. Clinical effectiveness and cost-utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation. *Health Technol Assess* 2003;**7**(9)
23 Woodroffe R, Yao GL, Meads C, *et al.* Clinical and cost-effectiveness of immunosuppressive regimens in renal transplantation. *Health Technol Assess* 2005;**9**(21)
24 Green C, Dinnes J, Hartwell D, *et al.* The clinical and cost-effectiveness of drotrecogin alfa (activated) (Xigris™) for the treatment of severe sepsis in adults: a systematic review and economic evaluation. *Health Technol Assess* 2005;**9**(11)

# Web appendix

## Methods

### Literature review methods

To identify papers reporting the results of models constructed to estimate the cost-effectiveness of AAA screening, the following search terms were used: 'Abdominal Aortic Aneurysm or AAA or Aneurysm'' and 'Screen∗' and 'Survival or Life Years or Quality Adjusted Life Years or Life Expectancy or QALYs' and 'Model∗'. The following bibliographic databases were searched in February 2005 – *Medline, EMBASE, EconLit, CINAHL, CAB Health, British Nursing Index (BNI), Biological Abstracts, MathSci, NHS Economic Evaluation Database (NHS EED), the Database of Abstracts of Reviews of Effects (DARE), the Office of Health Economics Health Economic Evaluations Database (OHE HEED), and the Health Technology Assessment (HTA) Report series*. All fields within databases were searched for papers or reports published up to and including 2004. No restrictions were placed upon publication language. Figure 1 shows the flow of publications.

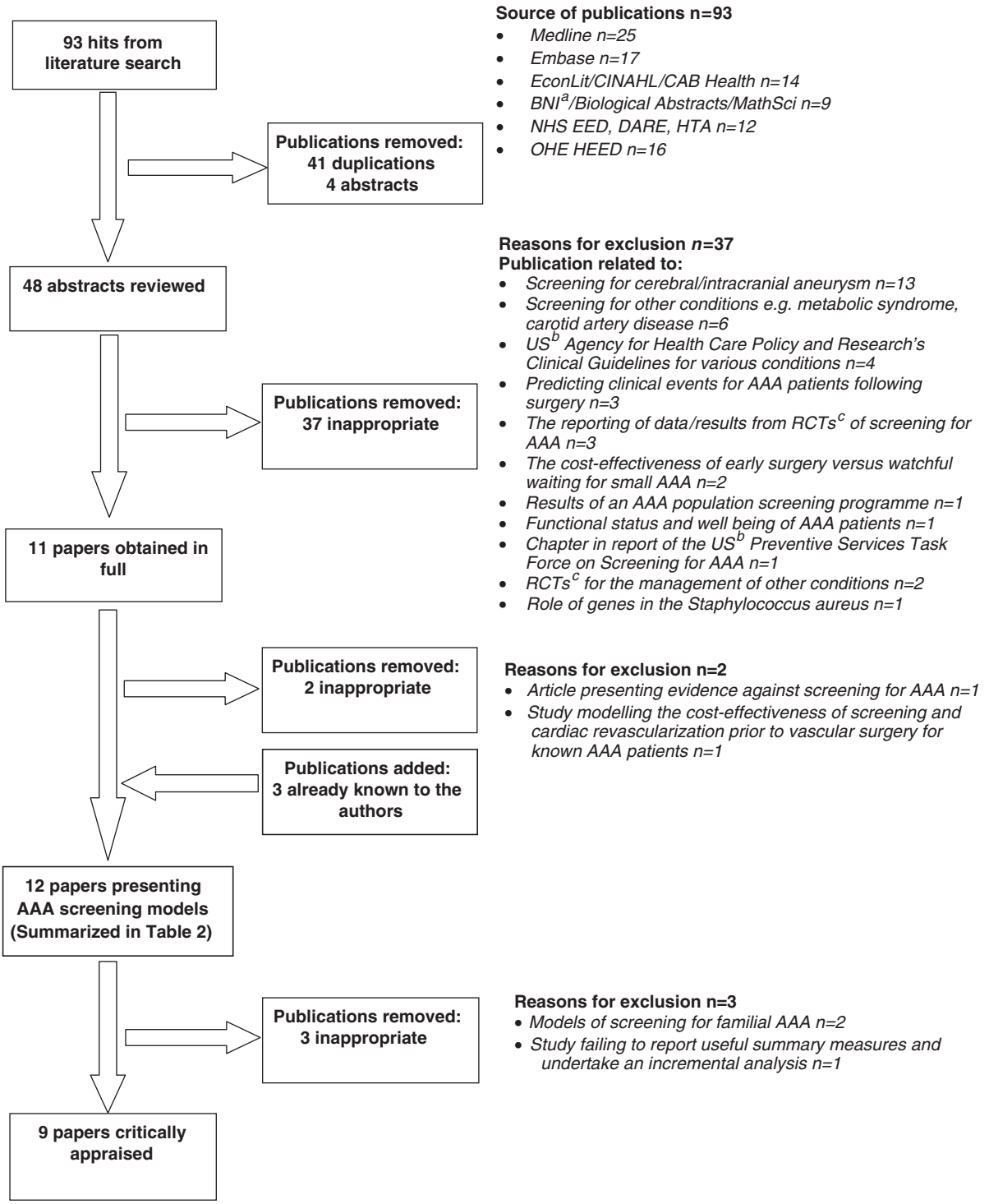
Papers identified as reporting health economic models were reviewed and results documented. To facilitate a comparison of results, all costs were inflated to 2003/2004 UK £ Sterling. For non-UK studies identified, costs were inflated to 2003/04 prices using country-specific inflation indices[1] before being converted to UK £ Sterling using published Purchasing Power Parities (PPPs).[2] Ideally, PPPs specific to AAA screening and treatment would have been used – such technology-specific indices are known to accurately account for issues of resource-use mix and price variation,[3] but these were not available. However, given that only western countries feature in the review, it is likely that the potential for resource-use mix to differ substantially in the provision of AAA screening and treatment is small.

## References

1 Organisation for Economic Co-operation and Development. *OECD Health Data 2003*. Paris, France: Organisation for Economic Co-operation and Development, 2003

2 Organisation for Economic Co-operation and Development. *OECD Health Data 2004* 1st edn. Comparative Analysis of 30 Countries Paris, France: Organisation for Economic Co-operation and Development, 2004

3 Wordsworth S, Ludbrook A. Comparing costing results in across country economic evaluations: the use of technology specific purchasing power parities. *Health Econ* 2005;**14**: 93–9




[a]BNI British Nursing Index
[b]US United States
[c]RCTs Randomized Controlled Trial

**Figure 1** Flow of publications identified

**Appendix Table 1** Structural elements of the nine cost-effectiveness models, ordered by date of publication

| | **Model** | | | | |
|---|---|---|---|---|---|
| Reference no. | 19<br>Bengtsson H, *et al*. | 5<br>Russell J | 18<br>Collin J | 14<br>Mason J, *et al*. | 15<br>Frame, *et al*. |
| Strategies | A US screen for men aged 60 with re-screen at 67 and 74 for those with negative results is compared with no screen | A US screen for males at age 60, then again at 65 and 70 is compared with no screen | A single US screen for men aged 65 is compared to no screen | A single US screen for men aged 70 is compared to no screen | A single US screen for men between 60 and 79 is compared with no screen where emergency and elective AAA treatment is permitted |
| Model type and structure | Described as a mathematical model. Appears to be based on principles of Markov modelling and is structured to reflect the underlying disease process and impact of screening | No model framework has been constructed | No model framework has been constructed | Model type not documented but appears to be based on the principles of Markov modelling and is structured to reflect the underlying disease process and impact of screening | Described as a computer spreadsheet model. Appears to be based on the principles of Markov modelling and is structured to reflect the underlying disease process and impact of screening |
| | Model appears to be used only for screening arm | | | Only patients for whom screening alters management are included in the model | Model is used for both screening and no screening arms |
| Disease states and cycle length | Health states include no AAA, AAA 29–39 mm, AAA > 39 mm | NA – no model constructed | NA – no model constructed | Health states include AAA 35–50 mm undetected, AAA ⩾ 50 mm undetected, AAA 35–50 mm detected, AAA ⩾ 50 mm detected | Health states include no AAA, AAA < 4 cm undetected, AAA < 4 cm detected, AAA > 4 cm undetected, AAA > 4 cm detected when < 4 cm and not examined since, AAA > 4 cm detected, death |
| | Cycle length appears to be one year | | | Cycle length appears to be one year | Cycle length appears to be one year |
| Time horizon | 20 years | Lifetime | Lifetime | Lifetime | 20 years |
| Structural assumptions | | No mention of opportunistic detection and elective AAA repair in the absence of screening | | Opportunistic detection and elective AAA repair in the absence of screening is assumed zero | |
| | 100% attendance at screening is assumed | | | | 100% attendance at screening is assumed |
| | Ultrasound is assumed 100% sensitive and specific | Ultrasound is assumed 100% sensitive and specific | Ultrasound appears to have been assumed 100% sensitive and specific | Ultrasound is assumed 100% sensitive and specific | Ultrasound is assumed 100% sensitive and specific |

| | **Model** | | | |
|---|---|---|---|---|
| Reference no | 6<br>Law *et al*. | 13<br>St Leger *et al*. | 16<br>Lee *et al*. | 17<br>Boll *et al*. |
| Strategies | Two US screens for men, one at age 60 and one at age 70, is compared with no screen | A single US screen for men aged between 68 and 72 is compared with no screen | A single US screen for men aged 70 is compared with no screen | A single US screen for men aged 60–65 is compared with no screen |
| Model type and structure | No model framework has been constructed | Described as a computer spreadsheet model. Appears to be based on the principles of Markov modelling | A Markov model is used and is structured to reflect the underlying disease process and the impact of screening | A Markov model is used and is structured to reflect the underlying disease process and the impact of screening |

**Appendix Table 1** (Continued)

| | Model | | | |
|---|---|---|---|---|
| | | and is structured to reflect the underlying disease process and impact of screening | | |
| | | Model is used for both screening and no screening arms | Model is used for both screening and no screening arms | Model is used for both screening and no screening arms |
| Disease states and cycle length | NA – no model constructed | Health states include No AAA, nine 'categories' for different AAA sizes (all detected and undetected), death, and survivor | Health states include AAA < 3 cm, AAA 3–4 cm, AAA 4–5 cm, AAA > 5 cm (all detected and undetected), surgery survivor, and death | Health states include No AAA, unknown small AAA, follow-up small AAA, unknown large AAA repaired AAA, rejected large AAA, and death |
| | | Cycle length is one year | Cycle length appears to be one year | Cycle length is one year |
| Time horizon | 20 years | Five-year model, with 10 years of life expectancy assigned to lives saved by screening | Lifetime | Lifetime |
| Structural assumptions | No mention of opportunistic detection and elective AAA repair in the absence of screening | Opportunistic detection and elective AAA repair in the absence of screening is assumed zero | | |
| | 100% attendance at screening is assumed | | 100% attendance at screening is assumed | |
| | Ultrasound appears to have been assumed 100% sensitive and specific | Ultrasound is assumed 100% sensitive and specific | Ultrasound appears to have been assumed 100% sensitive and specific | Ultrasound appears to have been assumed 100% sensitive and specific |
| | For structural assumptions relating to rupture and emergency repair with screening, see Table 3 in the journal | | | |

**Appendix Table 2** Assessment of methods used to identify, incorporate, and examine uncertainty around mortality rate for elective AAA repair

| Study | Baseline estimate of elective AAA repair mortality rate | Detail provided of bibliographic databases searched to inform parameter value | Data source(s) and methods used to generate baseline parameter value | Sensitivity analysis |
|---|---|---|---|---|
| Bengtsson H, *et al.*[19] | 6% | No – data described as coming from 'own experience' and published literature | Four studies (1 Australian, 1 US, 1 Swedish, 1 Norweigan), each reporting operative mortality among a series of patients undergoing elective AAA repair. Rates reported range from 3.7% to 12%<br>Method used to synthesize data not described | One-way sensitivity analysis using rates of 3% and 12% |
| Russell J[5] | 2% | No | No source referenced | No sensitivity analysis undertaken |
| Collin J[18] | 3% mentioned in text. Inclusion in model is unclear | None searched – model input data reported to be based upon local experience | No description provided of how figure was generated | No sensitivity analysis undertaken |
| Mason J, *et al.*[14] | 5% | No | A letter published in a UK journal where rates quoted are from 1.4% to 10%<br>Method used to synthesize data not described | One-way sensitivity analysis using rates of 4% and 3% |
| Frame P, *et al.*[15] | 5% | Systematic review augmented by an additional search of Medline and manual searching of paper bibliographies | Figure is reported in a paper publishing results of the Medline searches. Sixteen studies appear to have informed this value | Two multivariate sensitivity analyses setting all model parameters simultaneously to |

**Appendix Table 2** (Continued)

| Study | Baseline estimate of elective AAA repair mortality rate | Detail provided of bibliographic databases searched to inform parameter value | Data source(s) and methods used to generate baseline parameter value | Sensitivity analysis |
|---|---|---|---|---|
| | | | | values most and least favourable to screening |
| | | | Method used to synthesize data not described | |
| Law M, *et al.*[6] | 5% | No | Three studies (1 UK, 2 US), each reporting operative mortality among a series of patients undergoing elective AAA repair. Rates reported range from 1.4% to 4.9% | No sensitivity analysis undertaken |
| | | | Methods used to synthesize data are not described | |
| St Leger A, *et al.*[13] | 5% | No – model input data reported to be gleaned from the literature | One UK study reporting operative mortality among a series of patients undergoing elective AAA repair | One-way sensitivity analysis using rates of 0%, 2%, 4%, 5%, 6%, 8% and 10%. Method of presenting results, however, does not conform with convention* |
| Lee T *et al.*[16] | 4% | No – data were informed by a retrospective literature review | No sources referenced – details available from authors on request | One-way sensitivity analysis undertaken but uncertainty around elective AAA repair mortality rate was not explored |
| | | | Figure used is reported to be a weighted average | |
| Boll A *et al.*[17] | 6.8% | Published data from a Medline search and manual searching of bibliographies from relevant papers | Two papers, of which one is a 'seminar' published in a UK journal. Rates reported are from 4% to 8%. The source of these estimates is the second paper which reports operative mortality among all patients with non-ruptured AAA undergoing elective repair in the Netherlands in 1990 | One-way sensitivity analysis undertaken but uncertainty around elective AAA repair mortality rate was not explored |

*Results expressed as an 'absolute' cost per QALY calculated by dividing total costs in the screening arm of the model by the additional life years gained from screening

**Appendix Table 3** Assessment of methods used to identify, incorporate and examine uncertainty around the cost of emergency AAA repair

| Study | Baseline estimate of cost of emergency AAA repair (£ UK 03/04) | Detail provided of bibliographic databases searched to inform parameter value | Data source(s) and methods used to generate baseline parameter value | Sensitivity analysis |
|---|---|---|---|---|
| Bengtsson H *et al.*[19] | £16,040 | None searched – costs were calculated locally | Estimate based on data from 10 patients undergoing emergency repair in a Swedish General Hospital. No description is given of the costing methods used or of what resource-use is included in the estimate | One-way sensitivity analysis doubling the baseline cost of emergency repair |
| Russell J[5] | Not stated. Assumption appears to be that elective and emergency repair cost the same at £5,558 | No | If the aforementioned assumption is correct, the cost-estimate is based on a value reported in a letter published in a UK journal. The value per se appears to be based on expert clinical opinion | No sensitivity analysis undertaken |
| Collin J[18] | £11,115 | None searched – model input data reported to be based on local experience | No description provided of costing methods used or resource use included | No sensitivity analysis undertaken |

**Appendix Table 3** (Continued)

| Study | Baseline estimate of cost of emergency AAA repair (£ UK 03/04) | Detail provided of bibliographic databases searched to inform parameter value | Data source(s) and methods used to generate baseline parameter value | Sensitivity analysis |
|---|---|---|---|---|
| Mason J *et al.*[14] | £5,212 | No | Reported to be a personal communication from a UK hospital | Two-way sensitivity analysis using an extra-contractual referral cost for a general surgical procedure for both emergency and elective repairs |
| Frame P *et al.*[15] | £44,037 | Systematic review augmented by an additional search of Medline and manual searching of paper bibliographies | A published US study which costed surgery for 12 patients undergoing emergency repair. No description is given of the costing methods used and it is unclear what resource-use is included in the estimate | Two multivariate sensitivity analyses setting all model parameters simultaneously to values most and least favourable to screening |
| Law M *et al.*[6] | £11,115 | No | The modelling study by Collin J[18] | No sensitivity analysis undertaken |
| St Leger A *et al.*[13] | | | One published study, which applied costs from a UK hospital to estimates of average resource use consumed during an emergency AAA. Resource-use items are listed and a micro-costing approaches used | Baseline results generated using both published and calculated cost estimates |
| | £5,668 (locally calculated cost) | No – cost was obtained from the literature, and in addition a figure calculated locally | Second estimate was based on data from 8 patients undergoing emergency repair in a UK hospital. No description is given of the costing methods used or what resource-use is included in the estimate | One-way sensitivity analysis doubling the baseline cost of emergency repair. Method of presenting results, however, does not conform with convention* |
| Lee T *et al.*[16] | £19,553 | No – cost was calculated locally for a previous study | One published study in which unit costs from a US hospital are applied to literature based estimates of average resource-use consumed during an AAA emergency repair | One-way sensitivity analysis undertaken but uncertainty around the cost of emergency AAA repair was not explored |
| | | Resource-use items are listed and a micro-costing approach was adopted | | |
| Boll A *et al.*[17] | £10,894 | No – costs reported to be based on a screening feasibility study | No description provided of costing methods used or resource-use included | One-way sensitivity analysis undertaken but uncertainty around the cost of emergency AAA repair was not explored |

*Results expressed as an 'absolute' cost per QALY calculated by dividing total costs in the screening arm of the model by the additional life years gained from screening

**Appendix Table 4** Assessment of methods used to identify, incorporate and examine uncertainty around utility levels assigned to life years modelled

| Study | Utility level assigned to life years modelled | Detail provided of bibliographic databases searched to inform parameter value(s) | Data source(s) and methods used to generate baseline parameter value (s) | Sensitivity analysis |
|---|---|---|---|---|
| Russell J[5] | Three months of disability (utility score = 0.9) for survivors of AAA repair<br>All remaining life-years modelled are assigned a utility level of 1 | No | A UK publication provided the utility reduction estimate applied following AAA repair<br>A report on breast cancer screening was used to justify the assumption that 1 life year = 1 QALY | No sensitivity analysis undertaken |

**Appendix Table 4** (Continued)

| Study | Utility level assigned to life years modelled | Detail provided of bibliographic databases searched to inform parameter value(s) | Data source(s) and methods used to generate baseline parameter value (s) | Sensitivity analysis |
|---|---|---|---|---|
| Collin J[18] | All life years modelled are assigned a utility level of 1 | No | No sources referenced | No sensitivity analysis undertaken |
| St Leger A, *et al.*[13] | All life years modelled are assigned a utility level of 1 | No | No sources referenced | One-way sensitivity analysis undertaken, but uncertainty around utility levels was not explored |
| Lee T, *et al.*[16] | Post elective AAA repair 47 days of reduced utility | No – reported only to have been derived from the literature | No sources referenced - details available from authors on request | One-way sensitivity analysis undertaken, but uncertainty around utility levels was not reported |
| | Post emergency AAA repair 52 days of reduced utility | | | |
| | For survivors of AAA repair, utility is dependent on whether any long-term complications occur. Utility levels are as follows: | | | |
| | No complications | 1 | | |
| | Dialysis – dependent renal failure | 0.68 | | |
| | Stroke | 0.40 | | |
| | Myocardial infarction | 0.88 | | |
| | Major amputation | 0.70 | | |