

Discrete event simulation for decision modelling in health care: lessons from abdominal aortic aneurysm screening

MJ Glover^{*1}, E Jones^{*2}, KL Masconi², MJ Sweeting², SG Thompson² and the SWAN collaborators

1. Health Economics Research Group, Brunel University London

2. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge
SWAN collaborators are listed at the end of the paper

* Joint first authors.

Correspondence to: Mr M J Glover, Brunel University London, Uxbridge, UB8 3PH (email: matthew.glover@brunel.ac.uk)

Abstract

Markov models are often used to evaluate the cost-effectiveness of new healthcare interventions, but they are sometimes not flexible enough to allow accurate modelling or investigation of alternative scenarios and policies. A Markov model previously demonstrated that a one-off invitation to screening for abdominal aortic aneurysm (AAA) for men aged 65 in the UK and subsequent follow-up of identified AAAs was likely to be highly cost-effective at thresholds commonly adopted in the UK (£20,000–£30,000 per quality adjusted life-year). However, new evidence has emerged and the decision problem has evolved to include the exploration of the circumstances under which AAA screening may be cost-effective, which the Markov model is not easily able to address. A new model to handle this more complex decision problem was needed and the case of AAA screening thus provides an illustration of the relative merits of Markov models and discrete event simulation (DES) models. An individual level DES model was built using the R programming language to reflect possible events and pathways of individuals invited to screening versus those not invited. The model was validated against key events and cost-effectiveness as observed in a large randomized trial. Different screening protocol scenarios were investigated to demonstrate the flexibility of the DES. The case of AAA screening highlights the benefits of DES, particularly in the context of screening studies.

Abstract word count: 221

Article word count: 4,398

INTRODUCTION

It is recognised that modelling will almost always constitute an essential component of an economic evaluation to inform decision making, to overcome the limitations of available randomized trial data¹. In screening studies, where much of the cost is upfront whilst benefits accrue over a long period of time, there is a particular need for modelling approaches that can contribute to long-term economic evaluations. The choice of modelling technique is at the discretion of the analyst and often reflects an implied trade-off between simplicity and realism in reflecting a disease's natural history, treatment and patient outcomes^{2,3}. Markov models have been widely used since they provide a simple mechanism to estimate the long-term costs and effects of healthcare interventions⁴. Discrete event simulation (DES) is a less common alternative that avoids the use of states and fixed cycle lengths, and instead models events at the individual level. Some have suggested that DES should always be preferred, while others have highlighted particular circumstances where DES should be favoured^{5,6}. The case of abdominal aortic aneurysm (AAA) screening offers an illustration of the relative merits of these techniques.

An AAA is commonly defined as an aortic diameter $\geq 3.0\text{cm}$. A long-term Markov model demonstrated that offering population screening for AAA to men aged 65 in the UK was likely to be highly cost-effective⁷. This model was largely populated using data from the four-year follow-up of the randomized Multicentre Aneurysm Screening Study (MASS)⁸ and adopted the same screening methods, surveillance intervals (for 3.0-5.4cm AAAs) and AAA diameter threshold (5.5cm) for referral for elective surgical intervention as in MASS. The MASS trial of 67,800 men aged 65-74 showed that an invitation to a one-off ultrasound scan and surveillance or referral for elective surgical intervention of identified AAAs was effective in reducing AAA-related mortality, initially at 4-year follow-up⁸ and subsequently at longer term follow-up⁹. Results from the modelling and the MASS trial formed a large part of the evidence base supporting the phased implementation from 2009 of the NHS AAA Screening Programme (NAAASP) in England, with full coverage across the UK by the end of 2013.

Research on the clinical and cost-effectiveness of AAA screening has evolved since the first Markov modelling was performed as new data and evidence have emerged.

Initial observational data from NAAASP suggested that the current prevalence of AAAs is substantially lower than observed in MASS (1.6% vs 4.9%). The Markov model (MM) was updated to reflect this lower prevalence as well as changes in costs, the increased use of endovascular surgical techniques, meta-analysed data on growth and rupture rates^{10,11,11} and longer-term MASS follow-up¹². The results suggested screening is still likely to be highly cost-effective, with a long-term incremental cost-effectiveness ratio of £7,370 (95% CI £5,467 to £9,443) per Quality Adjusted Life Year (QALY). Other studies based on populations in Denmark¹³ and Sweden¹⁴, with similar AAA prevalence, support this conclusion.

However, different programmes and randomized trials have adopted diverse surveillance intervals, with little consensus on optimal intervals¹⁵. More substantial surveillance data from the RESCAN project was incorporated into an adapted MM to investigate different surveillance intervals and suggested that lengthening the time between rescans for men with the smallest aneurysms could be done at acceptable clinical risk¹⁰ and would be a cost-effective strategy¹¹. Some of the protocols around screening have also come under scrutiny. For example, the definition of an AAA as an aortic diameter $\geq 3.0\text{cm}$ is somewhat arbitrary: there is evidence that many individuals with screen-detected sub-aneurysmal aortic dilation (2.5-2.9cm) will progress to AAA within 10 years¹⁶. The implications for screening remain unclear. There have also been some suggestions that the surgical threshold itself should be altered¹⁷.

In light of these findings, the decision problem no longer relates only to “screening” versus “no screening” for elderly men, but has evolved to include the circumstances under which screening may be cost-effective¹⁸⁻²¹. Modelling allows these questions to be addressed without conducting costly primary research and could extend to varying a number of fixed parameters (e.g. surveillance intervals, the AAA diameter threshold for referral for elective surgery, screening of women, targeted screening based on patient characteristics). However, MMs can be inflexible. This inflexibility was demonstrated by the extensive re-programming needed to build tunnel states in the MM when the model was adapted to assess different surveillance intervals¹¹. Such analyses are important for existing programmes looking to improve their performance or extend population coverage, as well as for other countries considering implementation. Therefore, a model better able to handle this decision problem is

required. This paper describes: a) the development and validation of a DES model to estimate the clinical and cost-effectiveness of AAA screening; b) the use of the DES to explore the cost-effectiveness of screening under various scenarios, which was not possible with the original MM.

METHODS

Development of a simulation model

A DES was implemented using the freely available statistical programming language R and based on the original MM⁷. The original MM defined a number of health states that related to AAA identification, aortic diameter (<3.0cm, 3.0-4.4cm, 4.5-5.4cm, ≥5.5cm) and associated events (rupture, surgical consultations, elective and emergency AAA repair, death). A set of transition probabilities determined movements between health states for the two populations (invited to screening and not invited). The MM operates at the cohort level: events, mean costs and QALYs are calculated from the proportions of the cohort that inhabit the different health states in each 3-month cycle. In contrast, the DES functions at an individual level, simulating sequences of events that occur as a continuous process over time and calculating the associated mean costs and QALYs. It allows individual patient heterogeneity to be characterised and accounts for events as they occur, removing the need for any assumptions relating to averaging costs and/or outcomes across cycles.

An event scheduling approach

Full details of the DES are available in the SWAN project National Institute for Health Research Health Technology Assessment monograph. The DES adopts an event-scheduling approach by generating a sequence of events for each individual, using a list of events that are “scheduled” for the future (future events list; FEL). The DES has an explicit simulation clock and chooses the event that has the earliest sampled time, and records it in the individual’s sequence of events. It then schedules, reschedules, or cancels other scheduled events as necessary, updating the FEL (for example, if a surveillance rescan finds that an individual’s aortic diameter is above the threshold for elective surgery, then a consultation is scheduled). This process is repeated until

death or censoring (dependent on model time horizon). The possible sequences of events are shown in Figure 1.

Each individual is assigned an aortic diameter, drawn from a population distribution, and a latent parameter describing the growth rate of their aorta over time. Details of the aortic growth model are given in the Supplementary Material. Non-AAA death and AAA rupture events are scheduled in the future, and if the individual is in the “invited” group then an invitation is also scheduled. If the individual is in the “non-invited” group, then an “incidental detection” event is scheduled. The time to AAA rupture is dependent on the individual’s initial aortic diameter and their latent growth rate. In most instances, the scheduled AAA rupture time will be so far in the future that there is no chance of the event occurring. The “incidental detection” event is scheduled to occur only after the time at which an individual’s aortic diameter reaches the diagnosis threshold (e.g. 3.0cm).

The DES simulates people in pairs, like identical twins, one of whom is in the invited group and one in the non-invited group. The twins have certain characteristics in common: they have the same times of non-AAA death and AAA rupture in their FELs, the same initial aortic diameter and growth rate, and the same values of certain parameters such as indicators (binary variables) for whether they would be contraindicated for surgery and whether they would survive emergency surgery. Differences in costs and outcomes result from the different events that are scheduled due to involvement or otherwise in the screening programme.

Joint continuous AAA growth and rupture model

A major difference between the DES and the MM is that the DES uses a joint continuous-time model for aortic growth and rupture¹¹ rather than defining four AAA size states. Additionally, when an individual’s aorta is scanned, the measurement is generated by calculating the diameter according to this model and adding measurement error, which is specific to the type of scan used (i.e. ultrasound or computed tomography [CT]). Further alterations relating to the move from fixed cycles to a continuous process were made. For example, surgical waiting time was separated into two periods: the time from discovery to consultation (71 days) and from

consultation to surgery (59 days)²². In the MM, it was assumed that this total waiting time could be considered as a 3-month cycle. The DES would enable these periods to be easily changed if appropriate.

A hierarchy of functions

The R program for the DES is made up of a hierarchy of functions or routines: 1) a PSA (probabilistic sensitivity analysis), which consists of running the main analysis a repeated number of times; 2) the main individual patient simulation analysis, which consists simulating and analyzing multiple pairs of individuals; 3) the function to process one pair; and 4) the function to generate a sequence of events for an individual. These functions are shown in Figure 2.

The DES involves a large number of parameters. These can be classified into several sets: global fixed parameters, global uncertain parameters, and parameters that are specific to an individual or a pair of twins (“global” refers to population parameters and “uncertain” means that a parameter follows a random distribution). Like the functions, these sets form a hierarchy. For example, in a PSA, a beta distribution is used to generate the probability that an individual will die following emergency surgery, in the event that they have emergency surgery. The parameters of the beta distribution are global fixed parameters and the probability is a global uncertain parameter. In the main analysis, when a pair of twins is created, the probability is used as the parameter in a Bernoulli distribution to generate the indicator for the twin’s emergency surgery outcomes. The indicator is a variable specific to the pair of twins.

The hierarchy actually extends downwards by two more levels. When the DES generates an individual’s sequence of events it needs to record the intervention group (i.e. invited or not), which can be regarded as an individual-specific parameter, and when an aorta measurement is generated, a new and unique value of the measurement error is created, which can be regarded as an event-specific parameter. Figure 2 shows how the sets of parameters are passed from one function to another. The definitions of the parameters require judgement and depend on the nature of the input data. For instance, the costs of the scans and other events could be defined as

global fixed parameters, if their values are known with great certainty, or global uncertain parameters if they are not.

Cost-effectiveness analysis

The cost-effectiveness analysis consists of simulating a large number of individuals, calculating their life-years and costs, and calculating the mean life-years and costs over all the patients in the groups invited to screening and not invited. Given that the model outputs are driven by those individuals who have an AAA, these individuals were oversampled (and later calculations were adjusted to account for this), which reduces considerably the Monte Carlo error when estimating incremental effects and costs. PSA is conducted to account for uncertainty in the model parameters; repeated sets of values for the global uncertain parameters are generated, and the main analysis is run for each set of values. The mean incremental cost and effectiveness (i.e. QALYs), together with the incremental cost-effectiveness ratio (ICER) and incremental net monetary benefit (INMB), are calculated for each set, and the distribution of values is used to estimate the probability that the screening programme is cost-effective.

Programming practice

The DES is a moderately complicated computer program, and it was therefore necessary to follow basic principles of good programming to ensure that it would run correctly and be maintainable and usable in the future. For example, each function has a clearly defined single purpose that can easily be understood from its name, and priority was given to making the source-code simple, clear, and readable (by other people) rather than computationally fast. The DES was written using a mostly “functional” programming style: the basic building-blocks are functions, and functions do not modify things outside themselves but simply perform actions and then either display output or return relevant quantities (e.g. parameters, event-times). R is convenient for statistical and scientific programming, and it allows loops (or iterative processes) to be written to run in parallel (which is not the case with all programming languages). The DES is eminently suited to parallelisation, which speeds it up considerably.

Model validation

The original MM was validated against the MASS trial four-year follow-up to check the appropriateness of model outputs. The validation involved comparing the numbers of key events (e.g. AAA ruptures, number of elective operations) and mean costs and life-years, as observed in the trial, with model outputs based on a simulated population of the same size. The MM was able to replicate the observed data reasonably^{7,23}. For the DES, a similar process was carried out, again using the four-year MASS follow-up data. Costs and life-years were discounted at 6% and 1.5% per year, respectively, to be consistent with the original rates used in the four-year follow-up analysis.

Input parameters for the DES were derived from the MASS four-year follow up where possible, including non-AAA death rates to enable validation. Other adaptations were made to improve the DES model fit to the observed MASS outputs. The parameters for the aorta growth model were chosen such that at baseline there were the same proportions of individuals with aortic size <3.0cm, 3.0-4.4cm, 4.5-5.4cm and \geq 5.5cm as in MASS. Growth rates were based directly on those observed in the screen-detected MASS population with growth rates for those 2.0-2.9cm extrapolated from a fitted mixed model and growth rates set to zero for those <2.0cm at baseline (see Supplementary Material for more details). All aorta measurements that were performed by CT scan (at consultation only) were, on average, 0.24cm larger than an ultrasound scan, to account for CT scanning measuring outer-to-outer rather than inner-to-inner diameters¹¹. Individuals were censored at uniformly random times between 3 and 5.25 years, because the “four-year” follow-up of the MASS data actually had censoring times similar to this uniform distribution. Full details of input parameters and characterisation of uncertainty is detailed in Supplementary Table 1.

New model scenarios

After validation against the MASS trial four-year follow-up, parameter values in the DES were updated to reflect more contemporaneous estimates, full details of which are provided in Supplementary Table 2. National mortality statistics²⁴ were used for non-AAA death rates. The NAAASP baseline aortic diameter distribution was used in the aorta growth and rupture model, with growth and rupture rates based on RESCAN data²⁵. As before growth rates <3.0cm were extrapolated from a model or were set to

zero. Costs, attendance rate, and other parameters were updated as described in Glover et al (2014)¹², with QALYs estimated by applying population norm utility weights to life-years accrued. The model structure was further altered to allow for endovascular aneurysm repair (EVAR) as well as open repair, with a proportion of surgery by EVAR, which incurred a different cost and post-operative mortality rate compared to open surgery¹¹. The base case was run over a 30-year time horizon for 65-year-old men invited, or not invited, to screening, for 10^7 pairs of individuals. PSA based on 1000 runs with 500,000 pairs of individuals was used to characterize uncertainty in input parameters. Costs and QALYs were discounted at 3.5% per year.

Two modelling scenarios were explored, showing the flexibility of the DES to estimate the cost-effectiveness of screening under various protocols. The first built on an analysis previously performed using the MM, to identify more cost-effective surveillance intervals for men in the screening programme¹¹. The second allowed the inclusion of surveillance for men with sub-aneurysmal aortic diameter (2.5-2.9cm at first screen). Each of these different scenarios was compared to the existing programme in terms of costs and QALYs.

Scenario 1 – different surveillance intervals: Analysis performed using the MM suggested that lengthening surveillance intervals for the smallest identified AAAs would be cost-effective according to thresholds commonly adopted in the UK (£20,000-£30,000 per QALY)¹¹. However, the different surveillance strategies considered were limited to varying the time between monitoring for two AAA size groups (3.0-4.4cm and 4.5-5.4cm). Unlike the MM, the DES can be easily adapted to use any number of plausible AAA size cutoffs, or to consider differing surveillance intervals. Here the current NAAASP surveillance strategy of 1-year (3.0-4.4cm AAAs) and 3-month (4.5-5.4cm AAAs) intervals is compared to a strategy of 2-year (3.0-3.9cm AAAs), 1-year (4.0-4.4cm AAAs) and 3-month (4.5-5.4cm AAAs) intervals.

Scenario 2 – inclusion of sub-aneurysmal (2.5-2.9cm) aortas: The threshold definition of an AAA used in the DES was lowered from 3.0cm to 2.5cm. Individuals identified with an aortic diameter 2.5-2.9cm had surveillance scans scheduled at intervals of five years, with the intervals currently adopted by NAAASP maintained for AAAs between 3.0cm and 5.4cm.

RESULTS

DES model validation

The DES validated reasonably against the MASS four-year data. The DES broadly agreed with both the observed MASS four-year follow-up and the original MM in terms of differences in life-years and costs (Table 1). However, like the MM, the 4-year ICER for the DES was higher than in the MASS data. Table 2 shows the total numbers of events, as observed in MASS and estimated by the original MM and the DES. For most events, the numbers of events were similar, as the ratios of events in the DES to events in MASS show. The numbers of non-AAA deaths matched very closely, primarily because the MM and DES used non-AAA death rates from MASS and most individuals do not experience AAA rupture. As examples of cumulative events over time, Figure 3 shows the numbers of emergency operations in the non-invited group and AAA deaths in the invited group estimated by the DES, compared to the observed number in MASS four-year follow-up.

New model scenarios

The updated model after validation, using contemporaneous data sources, estimated a 30-year ICER of £6,352 (95% CI £5,059 to £8,808) per QALY (Table 3). This compares to a 30-year ICER of £7,370 produced by the MM¹². The 1,000 iterations on the cost-effectiveness plane and cost-effectiveness acceptability curve shown in Figure 4 demonstrate that a one-off invitation to AAA screening and subsequent follow-up of identified AAAs is highly likely to be cost-effective, with no iterations outside the cost-effective region.

Both the new scenarios estimated INMBs were positive compared to the existing screening programme at a willingness-to-pay threshold of £20,000 per QALY (Table 4). The longer surveillance interval for the smallest AAAs (scenario 1) resulted in a small cost saving, as those with 3.0-3.9cm AAAs are screened less often than in the existing programme. The longer surveillance interval was also associated with almost no change in QALYs. Extending the surveillance programme to those with 2.5-2.9cm aortic diameters (scenario 2) was associated with additional benefits, but greater costs. However, the model suggests that rescanning these individuals at 5-year

intervals could be cost-effective, with an INMB of £10 per man invited to screening. The estimate of the INMB for scenario 1 was positive, however the confidence interval from the PSA included zero; the probability of it being cost-effective compared to current strategy was 0.68. The INMB for scenario 2, with a confidence interval that excludes zero, suggests that a surveillance for those with aortic diameter between 2.5-2.9cm is a cost-effective strategy compared to the current strategy.

DISCUSSION

In assessing screening programmes, modelling techniques are particularly relevant given that a large proportion of costs are upfront, but benefits continue to accrue over time. The validated MM built by Kim et al⁷ demonstrated that AAA screening for men aged 65 in the UK was likely to be cost-effective, but the model was inflexible when trying to address questions around configuration and optimisation of screening. The creation of the DES has overcome these problems and the case of AAA screening highlights situations where DES may provide the most appropriate method to perform an economic evaluation, particularly when surveillance or rescreening is based on patient characteristics or risk markers (in this instance AAA size). However, it cannot be asserted that the decision around the conceptual model should have been different at the onset of the research. The evolution of the decision problem has necessitated the re-conceptualisation. Indeed, the MM served a valuable and timely purpose in showing the long-term cost-effectiveness of a one-off invitation to screening in the UK for men aged 65.

The DES built was based on the original MM, which itself was largely based on the MASS trial. This allowed the validation of the DES against observed MASS follow-up events and cost-effectiveness results. Overall the DES validation process was similar to that for the MM, and no major changes were necessary to produce a comparison of key outputs. Given some parameters in the model that are not based on data directly observed in MASS (e.g. incidental detection rate), further calibration to fit MASS observed data could have been undertaken, to better replicate the number of events and/ or cost-effectiveness results based on trial data. This type of calibration was performed to produce a better fit using the MM²⁴. However, when trying to validate a model against data from a study in this way, there is a risk of creating a cyclical process. If too much information from the study is used, then the model output might

match the data very well but the model may not predict well over a longer time horizon, rather like the issue of overfitting in statistical modelling. The differences between the DES and MM may be partly explained by the approach developed to handle aortic growth of those AAA <3.0cm at first screen. The performance of the DES in the validation gives some confidence in using the model to extrapolate over a longer term^{12,23}.

The general advantages of DES in health economic modelling have been extolled previously^{2,6,26-28}. They offer a decision modeller greater flexibility to adequately reflect clinical pathways, characterise baseline patient heterogeneity, allow event rates that change over time or depend on patient characteristics, and avoid the constraints of state transitions and fixed cycles imposed by a MM. In the case of AAA screening there are three particular characteristics that mean that a DES is superior; firstly being able to define the size of a AAA as a continuous variable, which also allows measurement error in the ultrasound observations; secondly, allowing heterogeneity in the AAA growth rates between different individuals, with uncertainty easily characterised, something that is difficult to recreate in an aggregate discrete-state MM, even by varying the transition probabilities over time; and thirdly the ease with which time-varying surveillance intervals and other changes to the screening programme can be defined and evaluated. There are generally perceived trade-offs between the simplicity of a MM and the complexity of a DES, in particular relating to: model build time, potential data requirements, and model run times. The latter will always be a consideration. However, the advantages of DES models start to outweigh other factors as the complexity of the decision problem and modelled pathway increases, especially if structural modifications and further data analysis are necessary to deal with different scenarios.

The results of the scenarios presented here would be difficult to replicate in a MM, constrained by the state transition approach. The problem of modelling different surveillance intervals was encountered in previous work on AAA screening and required the re-programming of 6 different incarnations of the MM¹¹, each containing a different number of 3-month tunnel states. For each surveillance interval that was different from the original screening strategy, tunnel states that allowed unobserved aortic growth and a related rupture rate in each cycle were necessary. Members of the

cohort in these unobserved tunnel states were then able to move back into observed states at each rescan. It would have been desirable to explore more combinations of screening intervals and associated AAA size cut-offs. However, the structural changes that are a necessity in an MM would have made this a time-consuming process, thus limiting the number of analyses that could be considered. Conversely, the DES can easily assess any combination of surveillance intervals. To change the interval for patients with a 3.0-3.9cm AAA from one to two years is trivial since the DES is programmed to allow the input of any chosen partition of the aortic size range with an associated screening interval for each part. This is only possible because an individual's AAA size is measured on a continuous scale. The problems of state transition are further demonstrated when trying to assess the cost-effectiveness of including sub-aneurysmal AAA in a screening programme. In the MM, a new AAA state (2.5-2.9cm with 5-yearly surveillance intervals) would need to be incorporated, with extensive reprogramming. In the DES, all that is necessary is to insert "2.5" in the list of surveillance thresholds and "5" in the list of intervals. In addition, the DES parameters can also be easily made dependent on individual-level covariates (e.g. age-dependent mortality rates after surgery).

The DES can be used to assess the cost-effectiveness of other policy relevant protocol changes including the surgical threshold, age at first screen and recalling all those screened normal at first screen after a period of time. The model has been parametrised as part of the SWAN²⁹ study and used to assess the likelihood of screening (with various different protocols) women being cost-effective.

The restructuring of the model as a DES was, as might be expected, a relatively complex undertaking. Nevertheless, coding in a language such as R enables greater clarity and transparency compared to software which is designed for simulation modelling. However, the computational requirements of the DES were extensive, given the number of individuals needed to reduce sampling variation to an acceptable level and characterising uncertainty through PSA. Run time was in the region of 24 hours to run the model with 500,000 patients and 1,000 PSA iterations, even with parallelisation and the use of a high-powered computer.

Acknowledgements

SWAN collaborative group: Simon G Thompson (Chief Investigator), Michael J Sweeting, Edmund Jones, Katya L Masconi (University of Cambridge); Janet T Powell, Pinar Ulug (Imperial College London); Matthew J Bown (University of Leicester); Matthew J Glover (Brunel University London).

The authors would like to thank colleagues at Brunel University London and in particular Professor Doug Coyle and Professor Martin Buxton for comments on versions of this manuscript. Any remaining deficiencies or inaccuracies are solely the responsibility of the authors.

Funding: This project was funded by the UK National Institute for Health Research (NIHR) Health Technology Appraisal (HTA) programme (project number 14/179/01). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, National Health Service or the Department of Health. Work done at the University of Cambridge was additionally funded by the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194), and the UK National Institute for Health Research (Cambridge Biomedical Research Centre)

Computer code: The R code developed for this project is available on request from the authors

REFERENCES

1. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: An unavoidable fact of life. *Health Econ.* 1997;6(3):217-227.
2. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ.* 2006;15(12):1295-1310.
3. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modeling practices: A suggested framework. *Value Health.* 1998;1(2):131-147.
4. Briggs A, Sculpher M. An introduction to markov modelling for economic evaluation. *Pharmacoeconomics.* 1998;13(4):397-409.
5. Karon J, Haji Ali Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics.* 2014;32(6):547-558.
6. Caro JJ, Moller J, Getsios D. Discrete event simulation: The preferred technique for health economic evaluations? *Value Health.* 2010;13(8):1056-1060.
7. Kim LG, Thompson SG, Briggs AH, Buxton MJ, Campbell HE. How cost-effective is screening for abdominal aortic aneurysms? *J Med Screen.* 2007;14(1):46-52.
8. Ashton HA, Buxton MJ, Day NE, et al. The multicentre aneurysm screening study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: A randomised controlled trial. *Lancet.* 2002;360(9345):1531-1539.
9. Thompson SG, Ashton HA, Gao L, Buxton MJ, Scott RA, Multicentre Aneurysm Screening Study (MASS) Group. Final follow-up of the multicentre aneurysm

screening study (MASS) randomized trial of abdominal aortic aneurysm screening.

Br J Surg. 2012;99(12):1649-1656.

10. RESCAN Collaborators, Bown MJ, Sweeting MJ, Brown LC, Powell JT, Thompson SG. Surveillance intervals for small abdominal aortic aneurysms: A meta-analysis. *JAMA.* 2013;309(8):806-813.

11. Thompson SG, Brown LC, Sweeting MJ, et al. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: Implications for surveillance intervals and their cost-effectiveness. *Health Technol Assess.* 2013;17(41):1-118.

12. Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the national health service abdominal aortic aneurysm screening programme in england. *Br J Surg.* 2014;101(8):976-982.

13. Lindholt JS, Juul S, Fasting H, Henneberg EW. Preliminary ten year results from a randomised single centre mass screening trial for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg.* 2006;32(6):608-614.

14. Svensjo S, Bjorck M, Wanhanen A. Current prevalence of abdominal aortic aneurysm in 70-year-old women. *Br J Surg.* 2013;100(3):367-372.

15. Stather PW, Dattani N, Bown MJ, Earnshaw JJ, Lees TA. International variations in AAA screening. *Eur J Vasc Endovasc Surg.* 2013;45(3):231-234.

16. Wild JB, Stather PW, Biancari F, et al. A multicentre observational study of the outcomes of screening detected sub-aneurysmal aortic dilatation. *Eur J Vasc Endovasc Surg.* 2013;45(2):128-134.

17. Paraskevas KI, Mikhailidis DP, Andrikopoulos V, Bessias N, Bell SP. Should the size threshold for elective abdominal aortic aneurysm repair be lowered in the endovascular era? yes. *Angiology*. 2010;61(7):617-619.
18. Powell JT, Thompson SG. Should the frequency of surveillance for small abdominal aortic aneurysms be reduced? *Eur J Vasc Endovasc Surg*. 2013;46(2):171-172.
19. Earnshaw JJ. Should the frequency of surveillance of a small abdominal aortic aneurysm be reduced? *Eur J Vasc Endovasc Surg*. 2013;46(2):174.
20. The International AAA Screening Group, Bjorck M, Bown MJ, et al. International update on screening for abdominal aortic aneurysms: Issues and opportunities. *Eur J Vasc Endovasc Surg*. 2015;49(2):113-115.
21. Davis M, Harris M, Earnshaw JJ. Implementation of the national health service abdominal aortic aneurysm screening program in england. *J Vasc Surg*. 2013;57(5):1440-1445.
22. Kim LG. A markov model for long-term cost-effectiveness modelling of screening for abdominal aortic aneurysms: Version 2. http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/A_Markov_model_for_MASS.pdf. Updated 2006. Accessed 9th July, 2016.
23. Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. *Health Econ*. 2010;19(1):43-55.
24. Office for National Statistics. Interim life tables, england and wales, 2009-2011. <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/o>

[ns/publications/re-reference-tables.html?edition=tcm%3A77-274529](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3990773/references/). Accessed 18 May, 2016.

25. Sweeting MJ, Thompson SG, Brown LC, Powell JT, RESCAN collaborators. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *Br J Surg.* 2012;99(5):655-665.
26. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ.* 2003;12(10):837-848.
27. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: Selecting the appropriate approach. *J Health Serv Res Policy.* 2004;9(2):110-118.
28. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: A report of the ISPOR-SMDM modeling good research practices task force--4. *Value Health.* 2012;15(6):821-827.
29. Screening women for abdominal aortic aneurysms (SWAN).
<https://www.journalslibrary.nihr.ac.uk/programmes/hta/1417901/#/> Accessed, 10th December 2016.

TABLES AND FIGURES

Table 1. Life-years and costs according to the four-year MASS follow-up, Markov model (Kim et al, 2007)⁷ and the DES.

	MASS observed	Markov Model	DES Model
No invitation group			
Life-years	3.816	3.905	3.753
Cost	£35.03	£32.74	£39.11
Invited group			
Life-years	3.819	3.907	3.754
Cost	£98.42	£98.32	£101.97
Difference			
Life-years	0.0022	0.0017	0.0015
Cost	£63.39	£65.58	£62.86
ICER	£28,400	£37,700	£42,137
(95% CI)	(£15,000, £146,000)	(£19,700, £147,000)	(£19,935, £3,277,596) †

Life-years discounted at 1.5% per year and costs at 6% per year

† Reported as uncertainty interval produced by 1000 probabilistic sensitivity analysis iterations (after assigning ICERs with negative incremental effects and positive costs to be infinite). Mean estimates from 1000 PSA iterations for the difference in life-years, costs, and the ICER were 0.0015, £62.91 and £46,032, respectively

Table 2. Numbers of key events observed in the MASS four-year follow-up, and as estimated by the Markov model (Kim et al, 2007)⁷ and the DES.

	MASS observed	Markov Model*	DES Model*	DES Model (% of MASS)
No invitation group				
Elective operation	100	83	98	98
Emergency operation	62	62	68	110
Rupture	138	141	154	112
Contraindicated for elective surgery	NA	14	16	NA
AAA death	113	109	120	106
Non-AAA death	3750	3724	3696	99
Invited group				
<i>Elective operation</i>				
Resulting from screen detection	295	282	330	112
Resulting from incidental detection	31	25	27	86
Emergency operation	28	34	30	106
Rupture	66	78	67	102
<i>Contraindicated for elective surgery</i>				
Resulting from screen detection	41	46	54	131
Resulting for incidental detection	NA	5	5	NA
AAA death	65	69	63	98
Non-AAA death	3694	3724	3700	100
Loss to recall follow-up	290	289	278	96

NA not available

* Estimated for a sample size of 33,961 participants in the control group and 33,839 in the invited group, as in MASS.

Table 3 Discrete event simulation: long term (30-year) cost–effectiveness of one-off invitation to AAA screening for 65-year-old men.

	DES Model
No invitation group	
Life-years	12.601
QALYs	9.681
Cost	£164
Invited group	
Life-years	12.611
QALYs	9.689
Cost	£213
Difference	
Life-years	0.01031
QALYs	0.00781
Cost	£50
ICER (QALYs) (95% CI)†	£6,352 (£5,059 to £8,808)

Life-years, QALYs and costs discounted at 3.5% per year

† Reported as uncertainty interval produced by 1000 probabilistic sensitivity analysis iterations. Mean estimates from 1000 PSA iterations for the difference in life-years, QALYs, costs, and the ICER were 0.01050, 0.00796, £50 and £6,388, respectively.

Table 4 Long term (30-year) cost-effectiveness. Scenario 1: surveillance intervals of 2 years (3.0-3.9cm AAAs), 1 year (4.0-4.4cm AAAs) and 3 months (4.5-5.4cm AAAs). Scenario 2: inclusion of sub-aneurysmal (2.5-2.9cm) AAAs in screening programme.

	Current strategy	Scenario 1	Scenario 2
Mean incremental QALYs	0.00781	0.00781	0.00860
Mean incremental cost	£49.61	£48.54	£55.17
Compared to current strategy:			
Mean incremental QALYs	n/a	0.00000	0.00080
Mean incremental cost	n/a	-£1.07	£5.56
ICER (QALYs) (95% CI)†	n/a	Dominant	£7,002 (4,615 to 12,233)
INMB* (95% CI)†	n/a	£0.99 (-2.03 to 3.35)	£10.33 (2.99 to 21.52)

Life-years, QALYs and costs discounted at 3.5% per year

* At a willingness-to-pay of £20,000 per QALY

† Reported as uncertainty interval produced by 1000 probabilistic sensitivity analysis iterations Mean estimates from 1000 PSA iterations for the difference in QALYs, costs, and the ICER for scenario 1 were 0.00000, £-1.08 and NA and for scenario 2 were 0.00080, £5.58 and £7,233, respectively.

Figure 1. Possible sequences of events in AAA screening DES model

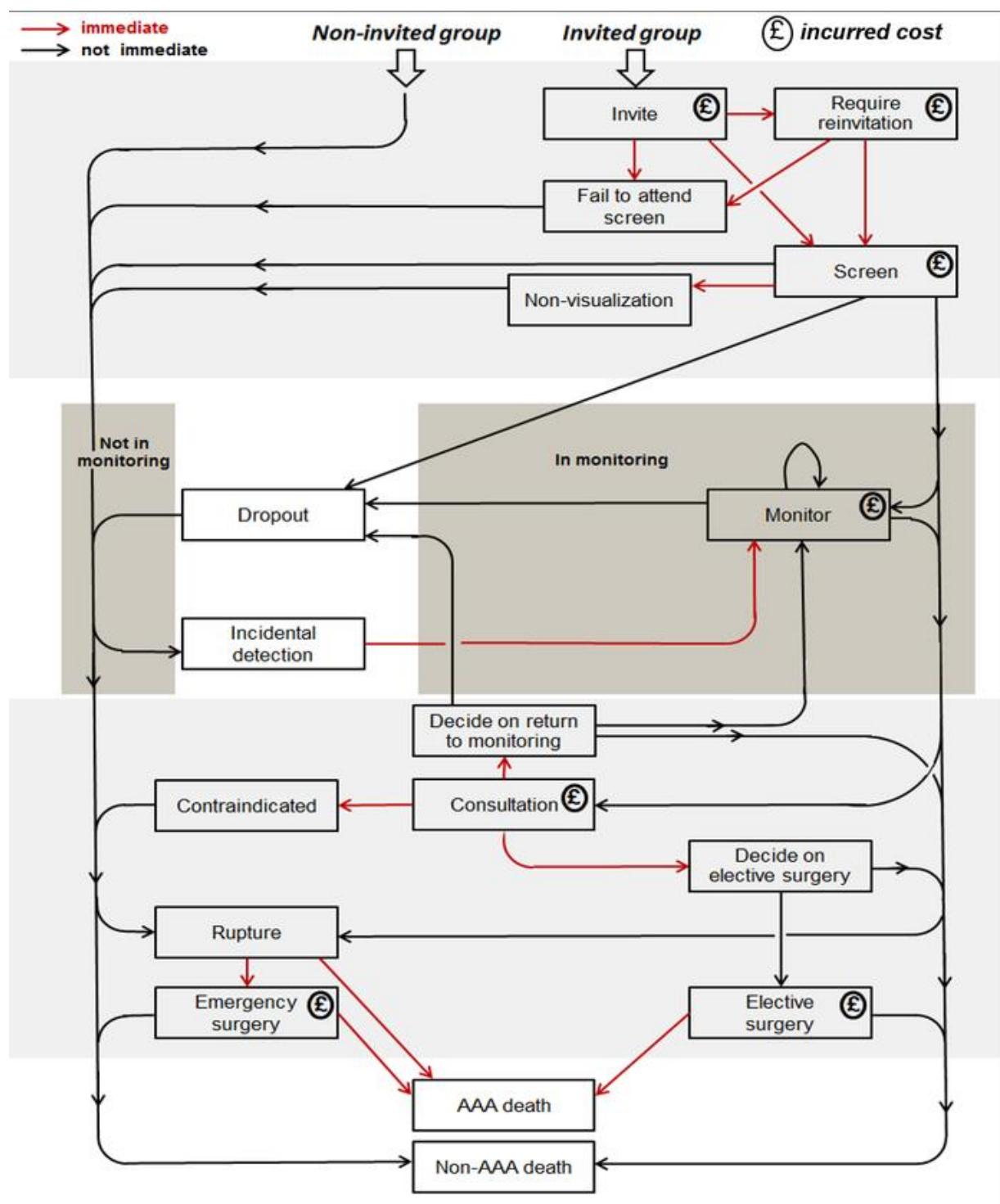


Figure 2. AAA screening DES model: hierarchy of functions.

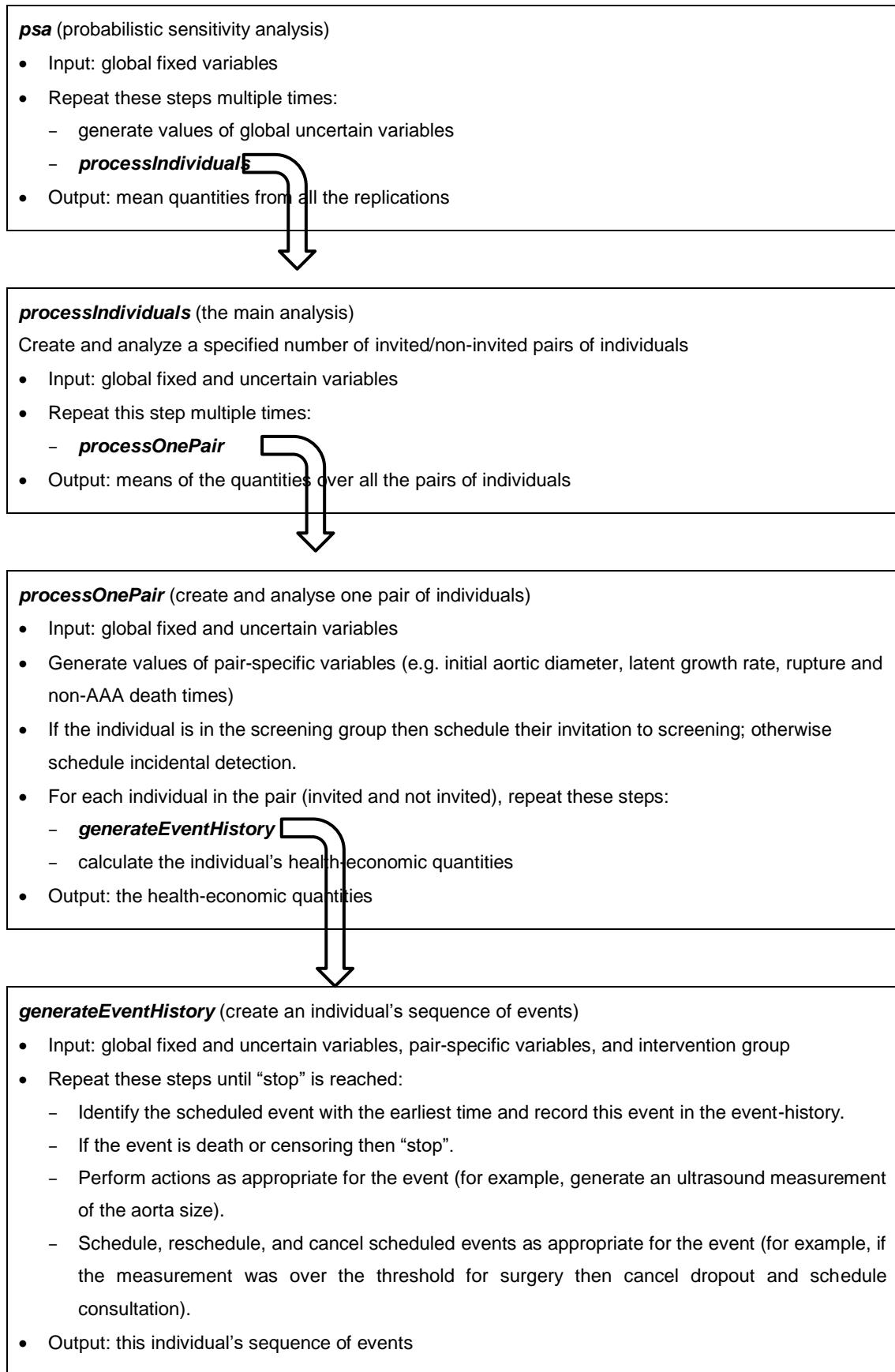


Figure 3. Cumulative numbers of events in the four-year MASS data and the DES for: a) Emergency operations in the non-invited arm, b) AAA deaths in the invited arm.

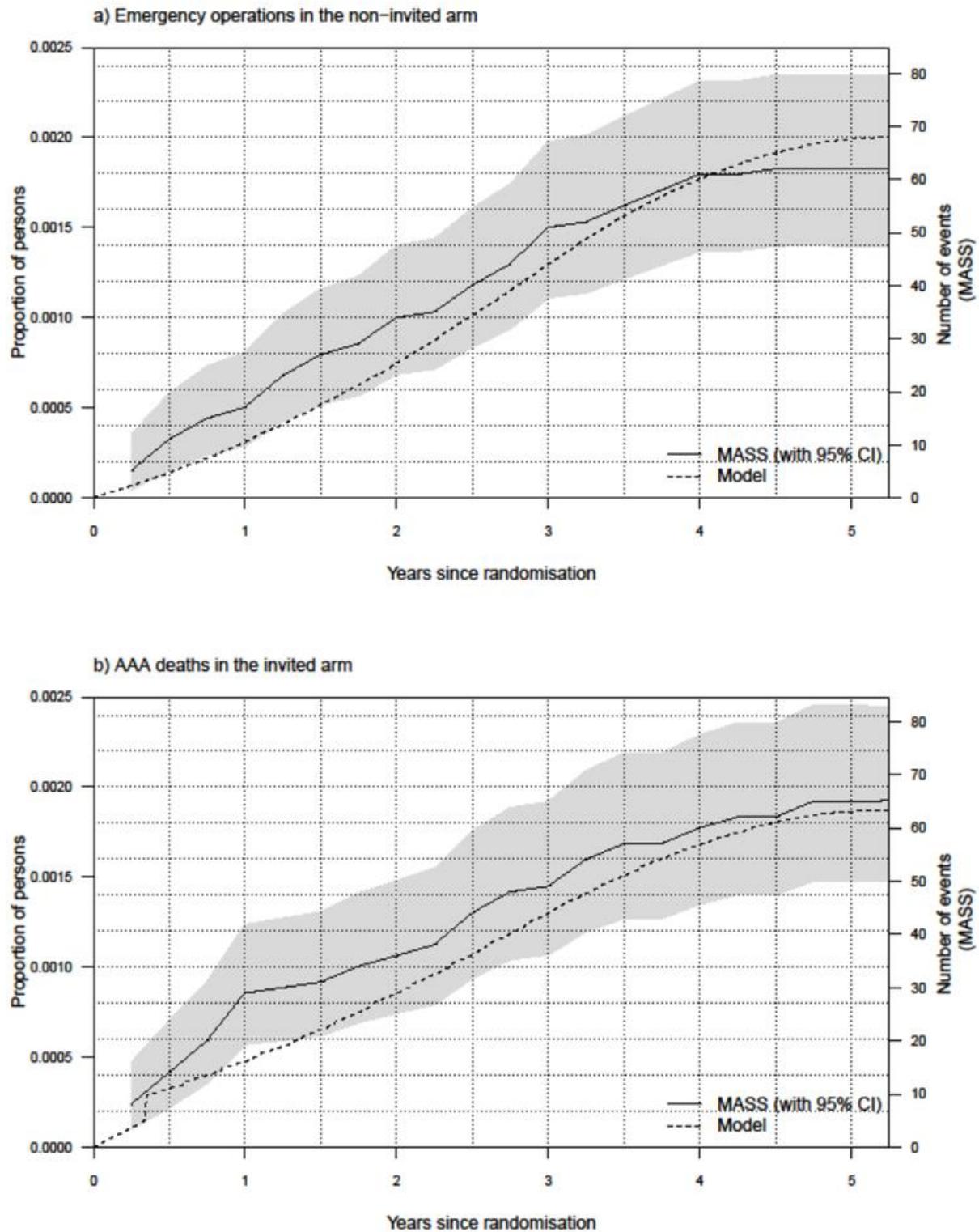


Figure 4. Long term (30-year) cost–effectiveness of one-off invitation to AAA screening: a) 1,000 probabilistic sensitivity analysis iterations (current NAAASP programme), b) Cost-effectiveness acceptability curve.

