Evaluation of a Stratified National Breast Screening Program in the United Kingdom: An Early Model-Based Cost-Effectiveness Analysis

DOI:
10.1016/j.jval.2017.04.012

Document Version
Accepted author manuscript

Evaluation of a Stratified National Breast Screening Programme in the United Kingdom: An early model-based cost-effectiveness analysis

Ewan Gray, PhD

Anna Donten, MSc

Nico Karssemeijer, PhD

Carla van Gils, PhD

D Gareth Evans, MD FRCP

Sue Astley

Katherine Payne*

Corresponding author*

Acknowledgements

The 'Adapting Breast Cancer Screening Strategy Using Personalised Risk Estimation' (ASSURE) project was funded from a Collaborative project grant within the FP7-HEALTH-2012-INNOVATION-1 call (Project number: 306088).

We acknowledge members of the ASSURE research team who contributed to the development of this analysis during useful discussions in project meetings.

Abstract (250 words)

Study objective: To identify the incremental costs and consequences of stratified national breast screening programmes (stratified-NBSP) and key drivers of relative cost-effectiveness.

Method: A decision-analytic model (discrete event simulation) was conceptualised to represent four stratified-NBSP (risk-1; risk-2; masking; masking and risk-1) compared with the current UK-NBSP and no screening. The model assumed a life-time horizon, the health service perspective to identify costs (£; 2015) and measured consequences in Quality Adjusted Life Years (QALYs). Multiple data sources were used including: systematic reviews of effectiveness and utility data; published studies reporting resource use and costs; cohort studies embedded in existing NBSP. Model parameter uncertainty was assessed using Probabilistic Sensitivity Analysis (PSA) and one-way sensitivity analysis.

Results: The base case analysis, supported by PSA, suggested risk-stratified-NBSP (risk-1; risk-2) were relatively cost-effective when compared with current UK-NBSP with incremental cost-effectiveness ratios (ICERs) of £16,689 per QALY and £23,924 per QALY, respectively. Stratified-NBSP including masking approaches (supplemental screening for women with higher breast density) was not a cost-effective alternative with ICERs of £212,947 per QALY (masking) and £75,254 per QALY (risk-1 and masking). When compared with no screening, all stratified-NBSP could be considered cost-effective. Key drivers of cost-effectiveness were: discount rate; natural history model parameters; mammographic sensitivity; and biopsy rates for recalled cases. A key assumption was that the risk model used in the stratification process was perfectly calibrated to the population.

Conclusion: This early model-based cost-effectiveness analysis provides indicative evidence for decision-makers to understand the key drivers of costs and QALYs for exemplar stratified-NBSP.

Introduction

National breast screening programmes (NBSPs) have emerged as important public health interventions that aim to reduce deaths from breast cancer through early detection [1]. NBSPs in different jurisdictions differ in terms of the age at which screening is first offered to women in the population (start of NBSP), the interval between screens (screening interval) and the age at which screening is stopped. In the UK, the current NBSP is targeted at women within the first three years of their 50th birthday until the age of 70 years with a three-yearly screening interval [2]. In some areas of the UK, the age range has been extended to women aged 47 to 49 years and 71 to 73 years as part of an age extension trial [3]. The current UK-NBSP is a standard programme with the same screening modality (mammography) offered at the same screening interval to all women regardless of their risk of developing breast cancer.

A new concept called 'stratified screening' also known as personalised screening, is being considered to replace the existing standard, or 'one size fits all' UK-NBSP, with the aim of improving the predictive value of cancer detection and, therefore, the relative cost-effectiveness of the programme[4]. Risks of breast cancer may vary across a wide range due to familial risk, mammographic density and modifiable risk factors. The potential for improved clinical and relative cost-effectiveness is achieved by modifying the screening protocol depending on an individual's characteristics such as breast cancer risk factors or the performance of the screening modality for that individual. The introduction, or any modification to, a NBSP has an opportunity cost. It is therefore important for decision makers deciding how to allocate finite budgets for screening programmes to understand the added value of any additions to or changes to a NBSP.

A substantial, but heterogeneous, economic evidence base has been developed to quantify the potential added value of NBSP. A systematic review, conducted in 2014, identified 71 economic evaluations of relevance to breast screening in a general population of women, of these, 52 were model-based evaluations [5]. There were three studies identified that conducted model-based analyses of a stratified screening strategy. Two of these studies were based in the USA [6,7] with no relevance to healthcare systems outside that setting. One study was UK-based [8] but provided no detail on the study perspective, time horizon, nature and source of model inputs or method of analysis, which meant it is not possible to critique the relevance and quality of the results. Given the lack of an existing evidence base, it was timely to design an early model-based cost effectiveness analysis (CEA) to identify the

potential impact of introducing stratified-NBSP in the UK-setting and key drivers of the relative cost-effectiveness of different types of stratified-NBSPs.

Method

An early model-based CEA was developed to address the key criteria described in Table 1 and reported in line with published criteria [9]. The concept of an early model-based economic evaluation is used in keeping with the definition offered by Annemans et al [10]. Using an early model-based economic evaluation is in keeping with the recommendation by Sculpher et al [11] to use an iterative approach to developing economic evidence to inform the introduction of new healthcare interventions.

<Table 1 here>

Interventions
Four potential approaches (hereafter (called: risk-1; risk-2; masking; masking and risk-1) to stratified-NBSP (see Table 1) were developed as part of an European-wide collaborative project called ASSURE (Adapting Breast Cancer Screening Strategy Using Personalised Risk Estimation) [4].

Comparators
The identified relevant comparator was the current UK-NBSP (see Table 1). No screening was also identified as a comparator of interest. A pragmatic approach was taken to define 'no screening' (see Table 1).

Model conceptualisation and structure
A systematic review of economic evaluations of breast screening programmes identified no relevant existing models that could be used without extensive modification [5]. A *de novo* model structure was conceptualised, in line with published recommendations [12] and developed with input from key clinical members in the ASSURE team (n=5) and external experts (n=15). The conceptualisation process identified that the model required three components to represent: the stratification approach; breast cancer natural history with screening; and the diagnosis and treatment process following a cancer detected by screening. A discrete event simulation model was used to represent these three components. Supplementary Appendix 1 shows the model structures and descriptions in detail. The model codes, created in R statistical package, are available on request.

Model input parameters

The input parameters, with key assumptions, are now described for each of the three model components together with the values used for resource use costs and outcomes, quantified using survival and quality-adjusted life years (QALYs).

The stratification process

Performance input parameters were required for each screening modality: mammography; mammography adjusted for masking; ultrasonography and magnetic resonance imaging.

Mammography

The sensitivity of mammography was defined as the conditional probability of a tumour being detected at a mammography event given the size of tumour. This model took account of latent cancers that exist at a screening round, which were not detected, and subsequently do not present in the following interval. To obtain an estimate of screening sensitivity consistent with the presence of latent cancers in the model the screening sensitivity as defined in Weedon-Fekjaer et al [13] was used. Screening sensitivity was estimated jointly with the natural history parameters to be consistent with the presence of latent cancers that were simulated in this model. Sensitivity of mammography conditional on tumour size was parameterised as shown in Equation 1:

$$S(X) = \frac{exp\left(\frac{X-\beta_2}{\beta_1}\right)}{1+exp\left(\frac{X-\beta_2}{\beta_1}\right)} \qquad \text{Equation 1}$$

Table 2 reports the definitions for the parameters used in Equations 1 to 12.
<Table 2>

Mammography and adjustment for masking

Masking was defined as the case in which a cancer was present but not detected at screening due to the view of the cancer being obscured in the images by other tissues [14]. In mammography, masking was expected to occur more frequently when there was high breast density or if particular textural patterns of the breast tissue were present. To quantify masking due to breast density it was necessary to rely on a comparison of screen-detected and interval breast cancer rates within different density groups. From such a comparison it was possible to estimate the sensitivity of screening mammography for each group by the method of counting the screen-detected cancers as true positives and the interval cancers as false negatives.

To calculate the Volpara Density Group (VDG) specific sensitivity (Sen_{VDG}; see Equation 2 and Table 2) of mammography for a tumour of a given size, the ratio of the odds of a true positive result for that VDG compared with the population average odds (OR_{VDG}; see Equation 3 and Table 2) was combined with the odds of true positive result given tumour size alone. The resultant value for odds was then converted back to a probability to give VDG-specific and tumour size-specific sensitivity. For simplicity, it was assumed that the relative sensitivities (i.e. odds ratios) between VDGs were equal across all tumour sizes.

$$OR_{VDG} = \frac{Sen_{VDG}/(1-Sen_{VDG})}{Sen_{average}/(1-Sen_{average})} \qquad \text{Equation 2}$$

$$Sen_{X,VDG} = \frac{Sen_X/(1-Sen_x)\times OR_{VDG}}{1+(Sen_X/(1-Sen_x)\times OR_{VDG})} \qquad \text{Equation 3}$$

Mammography recall rate (true positives and false positives)

The rate of recalls that result in biopsy (true positives) was taken from a previous economic evaluation [15]. The recall rate, for women in whom no cancer is present (false positives), was calculated by identifying the overall recall rate for the UK-NBSP from published programme statistics 2011-2012 [16]. Around 20% of recalls were cited to be true positives, which indicated the estimated recall rate, excluding true positives, was 3.2%.

Ultrasonography (US) and Magnetic Resonance Imaging (MRI)

Two supplemental screening modalities were relevant. Ultrasonography (US) supplemental screening, delivered either using hand-hand equipment (HHUS) or automated equipment (ABUS) was proposed for women with high breast density (VDG3 and VDG4). For women at high risk that also have high breast density Magnetic Resonance Imaging (MRI) was used as a supplemental screening technology.

It was necessary to assume that the only available published estimates of supplemental US and MRI screening sensitivity and specificity in this group were approximately equal to those for the relevant population (mammogram negative women of screening age).The estimate of US screening performance was taken from a published systematic review and meta-analysis [17]. This review only included studies in the 'high risk' population but was the only available source that provided a quantitative synthesis of sensitivity and specificity for US. For MRI, data from an ongoing trial in a high-risk population of women in this area, Vreeman et al (personal communication), was used to inform the MRI screening performance parameters in the model.

The same approach was taken to calculate the screening performance for US and MRI. Reported cancer detection rates from each source were used to calculate the odds ratio for detecting cancer with US, MRI and mammography compared with mammography alone. The estimated odds ratio was assumed to be constant across tumour size. Equation 4 (see Table 2), shows the case for MRI:

$$OR_{MRI} = \frac{c.d.r._{\cdot mammo,MRI}/(1000-c.d.r._{\cdot mammo,MRI})}{c.d.r._{\cdot mammo}/(1000-c.d.r._{\cdot mammo})} \qquad \text{Equation 4}$$

The cancer detection rate with mammography and MRI reported by Vreemann et al was 12.14 per 1000, while the cancer detection rate for mammography alone in this group was 4.2 per 1000 [17]. The estimated odds ratio was 2.91, which was then applied to the tumour size and breast density specific odds of a cancer being detected with mammography alone. These odds can then be converted back to probabilities for use in the simulation of individual screening events using the formula in equation 5 (see Table 2):

$$Sen_{X,VDG,MRI} = \frac{Sen_{X,VDG}/(1-Sen_{x,VDG})\times OR_{MRI}}{1+(Sen_{X,VDG}/(1-Sen_{x,VDG})\times OR_{MRI})} \qquad \text{Equation 5}$$

US and MRI recall rate

The recall rate for US was 98 per 1000 exams and for MRI it was 41 per 1000 exams [17]. It was assumed that the biopsy rate for recalls is the same as the current-NBSP, which was informed by the opinion of three experts (radiologists) in the ASSURE project [4].

Breast cancer natural history with screening

Breast cancer natural history was represented using a continuous time and tumour size growth model to allow variation in growth rates. The natural history of breast cancer was defined by estimating the incidence of breast cancer with screening and the growth of tumours once detected.

Breast cancer incidence

The occurrence of breast cancer for an individual was assumed equal to the life-time risk score of that individual, estimated using the Tyrer-Cuzick algorithm [18]. This assumption implies that the risk model used in the stratification process is perfectly calibrated to the population. The age of breast cancer incidence (malignant neoplasm of breast (ICD-9: C50) and carcinoma in-situ of breast (ICD-9: D05), conditional on life-time occurrence, was then

estimated for each individual based on Office of National Statistics (ONS) cancer registry data [19](see Supplementary Appendix 2).

Breast cancer growth

A continuous time model was used to estimate the growth of tumours of the breast (see Supplementary Appendix 3). Four candidate growth models ([13,20–22]) were identified from a systematic review of economic evaluations of NBSP [5] . Each identified growth model used a unique combination of parameters, which meant a formal quantitative synthesis was not appropriate, and the natural history model published by Weedon-Fekjaer et al (2008) [13] was judged to be the best available because of: the use of a continuous growth model; the high quality of the reporting and the relatively close match in location (Europe) and time period to the current UK setting. The natural history model parameterisation was described by two steps. The parameter estimates are listed in Table 3.

<Table 3 here>

Step 1. Equations 6 (see Table 2) and Equation 7 (see Table 2) show the logistic tumour growth function (Using tumour volume V mm^3, diameter s mm, time in years t and growth rate k, and assuming a spherical shape as is in Weedon-Fekjaer et al):

$$V(t) = \frac{V_{max}}{\left[1+\left(\left(\frac{V_{max}}{V_{cell}}\right)^{0.25}-1\right)e^{-0.25kt}\right]} \qquad \text{equation 6}$$

$$s(t) = 2\left(V(t)/\left(\frac{4}{3}\pi\right)\right)^{\frac{1}{3}} \qquad \text{equation 7}$$

Step 2. Equations 8 (see Table 2) and Equation 9 (see Table 2) show the extension to individual growth rates (mixed model):

$$V_i(t) = \frac{V_{max}}{\left[1+\left(\left(\frac{V_{max}}{V_{cell}}\right)^{0.25}-1\right)e^{-0.25k_it}\right]} \qquad \text{equation 8}$$

$$s_i(t) = 2\left(V_i(t)/\left(\frac{4}{3}\pi\right)\right)^{\frac{1}{3}} \qquad \text{equation 9}$$

Diagnosis and treatment process

Following a screen detected cancer, the model captured the diagnostic and subsequent treatment process. Three types of tumours for breast cancer were reflected in the model: invasive; none or micro-invasive; advanced.

Invasive tumours

For invasive cancers, the Nottingham Prognostic Index (NPI), a commonly used and validated classification system, was used to group the diagnosed tumours into three prognostic groups [23,24]. A systematic review was used to identify reported survival for NPI-defined sub-groups [see Appendix A4]. A meta-regression analysis showed there was substantial heterogeneity between the studies that was driven by the date in which the data were collected and a trend for improved survival over time, which implied it was more appropriate to select the most recent data to inform the probability of NPI group membership conditional on invasive tumour size category and survival for women diagnosed with breast cancer (see Table 2). Allocation of invasive cancer cases to NPI categories used the probability of NPI group membership conditional on tumour size category, as reported in Kollias et al 1999 [25], which was the only study identified reporting the required cross-tabulation of size and NPI category. The required probabilities of NPI sub-group membership were calculated using the reported cross tabulation of size category by NPI category (see Supplementary Appendix 3)

None or micro-invasive tumours

Three simplifying assumptions were made to capture the impact of detecting none-invasive or micro-invasive tumours; defined as 'ductal carcinoma in-situ (DCIS)'. A vanishingly small proportion of DCIS-tumours will not be screen-detected and, therefore, it was assumed that only screen-detected cancers may be assigned to the DCIS category. The proportion of screen-detected DCIS cancers was assumed to be constant regardless of the screening interval. This assumption was supported by the proportions of DCIS in screen-detected cancers in the UK-NBSP (three-year interval; 20.3%) [26] compared with the Netherlands-NBSP (two-year interval; 20.9%) [27] being similar. Survival for DCIS diagnosed and treated patients was assumed to be the same as for the general population in line with an audit of UK screen-detected breast cancers [26]. On this basis, any screen-detected cancer was given a probability of 0.203 of being assigned to the DCIS category. DCIS cancer cases have the same all-cause survival as the general population.

Advanced tumours

A small proportion of all breast cancers will present at the advanced stage with distant metastases defined as being Stage IV in TMN classification system [28]. The probability of a

breast cancer of a given size presenting at an advanced stage was assumed not to be related to the type of screening modality or interval. The source for the probabilities of advanced breast cancer at diagnosis conditional on tumour size was taken from the NHS audit of screen-detected breast cancers (2013) (see Supplementary Appendix 3). Estimates of 10-year survival for patients with advanced breast cancer were obtained from a meta-analysis of registries in six countries [29].

Survival, invasive (non-advanced) breast cancer

For women without a diagnosis of breast cancer, survival was taken from published population life-tables [30] and the parameters of a Weibull survival distribution were estimated. Simulation of individual age of mortality was achieved by inverting the Weibull cumulative distribution function and taking a random draw from the uniform(0,1) distribution using equation 10 (see Table 2):

$$T_m = \left(\frac{\log(U)}{\lambda}\right)^{1/\upsilon} \qquad \text{equation 10}$$

The observed effect of data collection date on survival from the meta-regression [manuscript under review] meant the most appropriate estimate of survival for women with a diagnosis of breast cancer was the most up-to-date estimate (see Fong et al 2015) [31]. The parameters of four functional forms for the baseline hazard function were estimated in a regression based survival analysis: exponential, Weibull, log-normal and log-logistic. The exponential model was selected model based on the Akaike Information Criterion (AIC; a measure of model fit) and visual inspection of Cox-Snell residuals (see Supplementary Appendix 4). Estimated coefficients (Table 3) from the parametric survival model were used to simulate a survival time by inverting the survival function and use of a random number generator using equation 11 (see Table 2):

$$T_c = \frac{-log(U(0,1))}{\gamma} \qquad \text{equation 11}$$

Fong et al (2015) presented data for women aged 50-65 years, including both screen-detected and interval cancers, and it was necessary to age adjust these data for women older than 65 years of age. A further adjustment was made to account for lead time in screen-detected cancers by reversing the process of lead time bias correction as described in [32] to introduce lead time for screen-detected cancers. Mortality from breast cancer for screen-detected cancers was, therefore, calculated from the simulated time the cancer would have presented clinically rather than the time of screen detection. Standard all-cause mortality was applied in the period between screen detection and clinical presentation. This adjustment implied that an assumption was made that there was no important short-term

negative effect on mortality from treatment. It was further assumed that breast cancer did not affect survival beyond ten-years after clinical presentation (hazard rate returns to the population rate). Overall survival time post breast cancer diagnosis (T_o) was calculated using equation 12 (see Table 2):

$$if\ (T_m \leq T_d), T_o = T_m$$
$$else\ T_o = T_c + T_d \qquad \text{equation 12}$$

Quality adjusted life years

QALYs were used to capture the consequence of each screening programme. In accordance with standard practice, life-years were adjusted for average health-related quality-of-life at a given age [33]. Estimates for these age-specific average utility weights were taken from [34]. The multiplicative method was used to combine health state utility weights and age-specific average utility weights [33]. Utility weights were identified by updating a published systematic review for breast cancer health states [35]. An identical search strategy limited to the period January 2010-October 2015 yielded 11 additional studies. Consistent with the suggestions made by Peasgood et al, heterogeneity in the studies meant that meta-analysis of utility weights was inappropriate [35]. Therefore relevant utility weights were identified from studies that most closely represented the health states in the model structure. No studies were identified that defined breast cancer health states for specific NPI categories. Therefore, the selected utility weights (see table 2) were taken from Lidgren et al [36] (2007) were used for early disease and advanced (distant metastases) disease, for the first year following diagnosis and subsequent years. These selected utility weights were assumed to also account for the impact of disutility from treatment, which is in keeping with the original source for these data.

Resource Use and Costs

In accordance with the assumed healthcare system perspective, resource use and associated costs accruing to the health services were used as model input parameters (see Table 2). Initial treatment and follow-up healthcare costs were included. Costs associated with treatment for breast cancer cases of DCIS, NPI-categories 1 to 3 and advanced cancer, were taken from a published study [24] [37] [15]. These estimates from 1992 were inflated to 2015 prices using the retail price index produced by the Office of National Statistics[38]. Supplementary imaging (US and MRI) costs were taken from the NHS schedule of reference costs (2013/14) from the categories: diagnostic whole breast ultrasound (no complications), and diagnostic breast MRI (no complications). Mammography costs were sourced from [15] and reflected estimates from a screening programme. An estimate of the cost of

administering risk and breast density based stratification was made based on experience from the PROCAS study [5]. The average cost per women was estimated as £10.57 (see Supplementary appendix 1 for further details).

Data analysis

The base case analysis calculated the total costs and QALYs for a sample of 100 million women over a life time from the relevant age (in years) reflecting the start of: each of the four specified stratified breast screening programmes; current UK-NBSP; no screening. Supplementary appendix 5 shows how using a sample of 100 million women should be sufficient to be confident that the model had sufficiently converged.

Incremental analysis was performed by comparing each stratified NBSP with (i) current NBSP and (ii) no screening. In addition, a full incremental analysis was performed. All costs and QALYs were discounted at a rate of 3.5%.

One-way sensitivity analyses were used to explore the impact of selected input parameters (see Supplementary Appendix 6). In addition, NICE recommend that a relevant sensitivity analysis for interventions such as screening with long-term outcomes is to apply a 1.5% discount rate for health outcomes and a 3.5% discount rate for costs [39]. In common with previously published economic evaluations in screening, a no discounting scenario was also estimated. Probabilistic sensitivity analysis (PSA) [40] was performed to quantify the effect of the joint uncertainty (see Supplementary Appendix 7) using a generalised additive model [41].

Results

Table 4 shows the results of the base case analysis for a: risk-based stratified-NBSP (using Risk-1 or Risk-2); masking-based stratified-NBSP (Masking); risk and masking-based stratified-NBSP (Risk-1 and Masking). The Risk-1 stratified-NBSP and Risk-2 stratified-NBSP were relatively cost-effective when compared with the current UK-NBSP. The Masking stratified-NBSP does not appear to be a cost-effective alternative when compared with the current UK-NBSP. Using an alternative discounting rate of 3.5% for costs and 1.5% for benefits, resulted in relatively lower estimated ICERs for all stratified-NBSP compared with the UK-NBSP. When compared with no screening, all screening programmes may be considered cost-effective. A full incremental analysis is available in Supplementary Appendix 7. This shows that masking and risk-1 and masking was dominated by the next alternative (current-NBSP and risk-1 stratified NBSP, respectively). The ICER for the remaining comparisons were: £23,197 per QALY for the current-NBSP compared with no screening;

£16,689 per QALY for risk-1 stratified NBSP compared with masking; £26,749 for risk-2 stratified NBSP compared with masking and risk-1 stratified NBSP.

<Table 4 here>

Sensitivity Analyses

To examine the decision between using the suggested stratified-NBSP and the current UK-NBSP a cost-effectiveness acceptability curve is presented in Figure 1 using the results of the PSA. Figure 2 shows the associated cost-effectiveness acceptability frontier, that suggests the current UK-NBSP would be selected as the preferred programme with a threshold of cost per QALY gained below £20,000 per QALY gained while the Risk-2 stratified NBSP would be chosen at higher thresholds of cost per QALY gained.

<Figure 1 here>
<Figure 2 here>

One-way sensitivity analysis (see Supplementary Appendix 6) showed that the reported total costs, total QALYs and ICERs were sensitive to: natural history parameter values (α_2 and mean tumour size at clinical detection) and screening performance of mammography (β_2). ICERs for stratified programmes were moderately sensitive to the cost of stratification although costs would need to be several times the base case value for ICERs to in increase beyond a threshold of £30,000 per QALY. In all alternative programmes total costs were sensitive to the treatment cost parameters however varying these parameters did not greatly change the ICERs compared with the base case. Estimates of total QALYs were sensitive to the utility weights for cancer states, varying utility weights moderately altered the ICERs of stratified programmes compared with the NBSP. The results were relatively insensitive (within the ranges tested) to: the probability of recall; costs of MRI; the relative sensitivity of mammography by VDG group; and US/MRI additional cancer detection rate.

Discussion

This study used an early model-based CEA to generate estimates of the relative costs and consequences of four example stratified-NBSPs compared with no screening and current practice in the UK-NBSP. The Risk-1 and Risk-2 stratified-NBSP compared to the current UK-NBSP were deemed to be a cost-effective use of healthcare resources relative to a threshold range of £20,000 to 30,000 per QALY gained. The ICERs for the current UK-NBSP compared with no screening were somewhat higher than previous analyses [15,42] but were very similar to the most recently published study [43]. Results were not directly comparable

to previous model-based analyses of stratified screening [6,7] due to differences in modelling strategy and also comparators.

The Masking stratified-NBSP was relatively the less cost-effective strategy. Combining the two stratification approaches using Risk-1 and Masking simultaneously resulted in modest QALY gain when compared with either Risk 1 or Masking stratified NBSP. The modest gains from masking based strategies could be due to increased over-diagnosis overwhelming the potential QALY gains from early detection of a tumour. Over-diagnosis is a commonly cited problem with NBSP [1]. Over-diagnosis suggests that NBSPs are too effective at detecting small, and slow growing, tumours that would not affect a woman's health within her lifetime if left undetected. Follow up procedures such as biopsies and treatment for such over-diagnosed cases are expensive and may cause harm [44].

The interpretation of the cost-effectiveness results for stratified breast screening was strongly influenced by the choice of discount rate. The choice of discount rate is not a simple technical question and the preferred discounting procedure for producing cost-effectiveness results for economic evaluations in health is a contested issue [39]. Decision makers should consider which discounting scenario best reflects the values and preferences of those for whom they are making a decision,

This early economic analysis was based on the best available data sourced from a combination of rapid reviews, systematic reviews and analysis of data from two key published prospective studies [31,45]. Key data gaps were: the relative sensitivity of mammography by density given the tumour size; the detection rate of supplemental ultrasound; recall rate and biopsy rate. Most importantly, the lack of randomised trials, or sufficiently long robust observational studies, meant that there were no direct estimates of the effect of supplemental screening modalities on mortality or other long-term outcomes. Robust, up-to-date, data on the cost of treating women with breast cancer were not available. This meant it was necessary to rely on estimates from a now dated study for the cost of treatment stratified by a prognostic indicator [37]. In addition, on the advice of clinical experts the implications of screening on use of different targeted treatment options based tumour HER2 or ER receptor status was not included in this model. These important uncertainties, due to the lack of robust data for several key parameters suggest that the results of this model-based CEA should be treated as indicative. The focus should be on the model structure itself and on the identified key drivers of relative cost-effectiveness. The most important drivers of cost-effectiveness after the discount rate were the natural history parameters, cost of stratification and mammographic sensitivity parameters and future research should be directed at improving the robustness of these data. Some one-way

sensitivity analysis results may appear inconsistent (US and MRI cancer detection rates) and this may be due to Monte Carlo error for alternatives where the differences between strategies were small or the result of non-linearity in the model.

Decision makers using the results of this DES model-based CEA must recognise the inherent limitations of mathematical models of disease natural history and screening that may introduce structural uncertainty. Using a DES, was in line with published models in cancer screening [46]. DES allowed the influence of individual patient-characteristics to be captured; the flexibility for cancer growth to be modelled as a continuous process; the use of prognostic categories to group treatment options. It may be that modelling choices, such how cancer growth rates can vary between individuals, were influential in driving the relative cost-effectiveness of a particular NBSP. No formal external validation or calibration of this early decision analytic model was conducted. External validation against more extensive clinical trial or observational data should be a goal of any future investigation of the cost-effectiveness of stratified-NBSP.

A key important assumption, in the absence of data to prove otherwise, was that the risk model used in the stratification process was perfectly calibrated to the population. This ‘structural’ uncertainty is not reflected results of PSA therefore users must exercise judgement when interpreting the results. A further limitation to be aware of is that the use of regression models within a PSA is a new and developing methodology therefore these results should perhaps be treated with some caution. Structural uncertainty may be best addressed by planning external validation studies in future research relating to all aspects of the economic model, including, the risk models used in stratification and the natural history models of breast cancer. External validation studies of the risk model to be used in a stratified-NBSP are essential if there is reason to believe calibration may be poor, which also requires consensus to be reached on the appropriate risk categories to use in practice. Previous experience in a research context suggests that embedding stratification in the existing NBSP is feasible [47] but no data exist on the effects of stratification on screening uptake and this is an important topic for further research.

Conclusions

This early model-based CEA presents indicative results that suggest a risk stratified-NBSP is potentially a cost-effective use of healthcare resources when compared to the current UK-NBSP. The proposed model structure will be a key resource as more data become available to support the introduction of stratified-NBSP such as the sensitivity and effectiveness of the

new screening modalities, the effect of risk communication strategies on NBSP uptake and the cost of newer treatments for breast cancer. The choice of discount rate will be crucial in interpreting the results. A pre-specified external validation analysis should be conducted alongside any more definitive economic evaluation.

References

[1] Marmot M, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M. The Benefits and Harms of Breast Cancer Screening: An Independent Review. London, UK: Cancer Research UK and the Department of Health (England); 2012.

[2] Breast screening: programme overview 2015. https://www.gov.uk/guidance/breast-screening-programme-overview#screeningage (accessed May 30, 2016).

[3] Breast screening: age extension trial 2015. https://www.gov.uk/government/publications/nhs-breast-screening-age-extension-trial (accessed May 30, 2016).

[4] ASSURE: Personalised Breast Screening 2014. http://www.assure-project.eu/ (accessed May 30, 2016).

[5] Evans G, Stavrinos P, Dawe S, Harvie M, Wilson M, Maxwell A, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and Family History Clinic: a dual cohort study. 2015.

[6] Tosteson ANA, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-Effectiveness of Digital Mammography Breast Cancer Screening. Ann Intern Med 2008;148:1. doi:10.7326/0003-4819-148-1-200801010-00002.

[7] Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing Mammography by Breast Density and Other Risk Factors for Breast Cancer : Analysis of Health Benefits and Cost-Effectiveness. Ann Intern Med 2011;155:10–21.

[8] Van Dyck W, Gassull D, Vértes G, Jain P, Palaniappan M, Schulthess D, et al. Unlocking the value of personalised healthcare in Europe—breast cancer stratification. Heal Policy Technol 2012;1:63–8. doi:10.1016/j.hlpt.2012.04.006.

[9] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMC Med 2013;11:80. doi:10.1186/1741-7015-11-80.

[10] Annemans L, Genesté B, Jolain B. Early Modelling for Assessing Health and Economic Outcomes of Drug Therapy. Value Heal 2000;3:427–34. doi:10.1046/j.1524-4733.2000.36007.x.

[11] Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy 1997;2:26–30. doi:10.1177/135581969700200107.

[12] Roberts M, Russell LB, Paltiel a D, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Med Decis Mak 2012;32:678–89. doi:10.1177/0272989X12454941.

[13] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Res 2008;10:R41. doi:10.1186/bcr2092.

[14] Harvey JA, Bovbjerg VE. Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk1. Radiology 2004;230:29–41. doi:10.1148/radiol.2301020870.

[15] Madan J, Rawdin A, Stevenson M, Tappenden P. A rapid-response economic evaluation of the UK NHS Cancer Reform Strategy breast cancer screening program extension via a plausible bounds approach. Value Health 2010;13:215–21. doi:10.1111/j.1524-4733.2009.00667.x.

[16] Patnick J. NHS Breast Screening Programme Annual Review 2012. 2012.

[17] Tice J, Ollendorf D, Lee J, Pearson S. The Comparative Clinical Effectiveness and Value of Supplemental Screening Tests Following Negative Mammography in Women with Dense Breast Tissue. 2013.

[18] Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 2004;23:1111–30. doi:10.1002/sim.1668.

[19] Office of National Statistics. Cancer Incidence and Mortality in the United Kingdom - 2008-10. 2012.

[20] Baker RD. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Manag Sci 1998;1:103–13.

[21] Van Oortmarssen GJ, Habbema JDF, Van Der Maas PJ, De Koning HJ, Collette HJ a., Verbeek ALM, et al. A model for breast cancer screening. Cancer 1990;66:1601–12. doi:10.1002/1097-0142(19901001)66:7<1601::AID-CNCR2820660727>3.0.CO;2-O.

[22] Tan KHX, Simonella L, Wee HL, Roellin A, Lim Y-W, Lim W-Y, et al. Quantifying the natural history of breast cancer. Br J Cancer 2013;109:2035–43. doi:10.1038/bjc.2013.471.

[23] Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361–6.

[24] Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat 1992;22:207–19. doi:10.1007/BF01840834.

[25] Kollias J, Murphy CA, Ellis IO, Elston CW, Robertson JFR, Blamey RW. The prognosis of small primary breast cancers. Eur J Cancer 1999;35:908–12.

[26] [UK NHSBSP and ABS Screening Audit Group]. NHS breast screening programme and association of breast surgery; An audit of screen detected breast cancers for the year of screening April 2013 to March 2014. 2015.

[27] Bluekens AMJ, Holland R, Karssemeijer N, Broeders MJM, den Heeten GJ. Comparison of Digital Screening Mammography and Screen-Film Mammography in the Early Detection of Clinically Relevant Cancers: A Multicenter Study. Radiology 2012;265:707–14. doi:10.1148/radiol.12111461.

[28] Sobin LH, Gospodarowicz MK (Mary K., Wittekind C (Christian), International Union against Cancer. TNM classification of malignant tumours. Wiley-Blackwell; 2009.

[29] Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007. Thorax 2013;68:551–64. doi:10.1136/thoraxjnl-2012-202297.

[30] Office of National Statistics. United Kingdom, National Life Tables, 1980-82 to 2011-13. 2014.

[31] Fong Y, Evans J, Brook D, Kenkre J, Jarvis P, Gower-Thomas K. The Nottingham Prognostic Index: five- and ten-year data for all-cause survival within a screened population. Ann R Coll Surg Engl 2015;97:137–9. doi:10.1308/003588414X14055925060514.

[32] Duffy SW, Nagtegaal ID, Wallis M, Cafferty FH, Houssami N, Warwick J, et al. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. Am J Epidemiol 2008;168:98–104. doi:10.1093/aje/kwn120.

[33] Ara R, Wailoo A. NICE DSU Technical Support Document 12: the use of health state utility values in decision models. … , UK Sch Heal Relat Res … 2011:1–30.

[34] Ara R, Brazier JE. Using Health State Utility Values from the General Population to Approximate Baselines in Decision Analytic Models when Condition-Specific Data are Not Available. Value Heal 2011;14:539–45. doi:10.1016/j.jval.2010.10.029.

[35] Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev Pharmacoecon Outcomes Res 2010;10:553–66. doi:10.1586/erp.10.65.

[36] Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Qual Life Res 2007;16:1073–81. doi:10.1007/s11136-007-9202-8.

[37] Johnston K. Modelling the future costs of breast screening. Eur J Cancer 2001;37:1752–8. doi:10.1016/S0959-8049(01)00202-7.

[38] Office of National Statistics. Consumer Price Indices - RPI indices: 1987 to 2015. 2015.

[39] Jonsson P, Robertson J. Briefing paper for methods review working party on discounting. 2013.

[40] Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339–47. doi:10.1002/hec.985.

[41] Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making 2014;34:311–26. doi:10.1177/0272989X13505910.

[42] Forrest P. Breast Cancer Screening, Report to Health Ministers of England, Wales, Scotland & Northern Ireland 1986.

[43] Pharoah PDP, Sewell B, Fitzsimmons D, Bennett HS. Cost effectiveness of the NHS breast screening programme : life table model. Br Med J 2013;2618. doi:10.1136/bmj.f2618.

[44] Alvarado M, Ozanne E, Esserman L. Overdiagnosis and overtreatment of breast cancer. Am Soc Clin … 2011.

[45] Evans DGR, Donnelly LS, Harkness EF, Astley SM, Stavrinos P, Dawe S, et al. Breast cancer risk feedback to women in the UK NHS breast screening population. Br J Cancer 2016;114:1045–52. doi:10.1038/bjc.2016.56.

[46] Pilgrim H, Tappenden P, Chilcott J, Bending M, Trueman P, Shorthouse A, et al. The costs and benefits of bowel cancer service developments using discrete event simulation. J Oper Res Soc 2008;60:1305–14. doi:10.1057/jors.2008.109.

[47] Evans D, Astley S, Stavrinos P, Harkness E, Donnelly L, Dawe S, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and Family History Clinic: A dual cohort study. 2016.

Abstract (250 words)

Study objective: To identify the incremental costs and consequences of stratified national breast screening programmes (stratified-NBSP) and key drivers of relative cost-effectiveness.

Method: A decision-analytic model (discrete event simulation) was conceptualised to represent four stratified-NBSP (risk-1; risk-2; masking; masking and risk-1) compared with the current UK-NBSP and no screening. The model assumed a life-time horizon, the health service perspective to identify costs (£; 2015) and measured consequences in Quality Adjusted Life Years (QALYs). Multiple data sources were used including: systematic reviews of effectiveness and utility data; published studies reporting resource use and costs; cohort studies embedded in existing NBSP. Model parameter uncertainty was assessed using Probabilistic Sensitivity Analysis (PSA) and one-way sensitivity analysis.

Results: The base case analysis, supported by PSA, suggested risk-stratified-NBSP (risk-1; risk-2) were relatively cost-effective when compared with current UK-NBSP with incremental cost-effectiveness ratios (ICERs) of £16,689 per QALY and £23,924 per QALY, respectively. Stratified-NBSP including masking approaches (supplemental screening for women with higher breast density) was not a cost-effective alternative with ICERs of £212,947 per QALY (masking) and £75,254 per QALY (risk-1 and masking). When compared with no screening, all stratified-NBSP could be considered cost-effective. Key drivers of cost-effectiveness were: discount rate; natural history model parameters; mammographic sensitivity; and biopsy rates for recalled cases. A key assumption was that the risk model used in the stratification process was perfectly calibrated to the population.

Conclusion: This early model-based cost-effectiveness analysis provides indicative evidence for decision-makers to understand the key drivers of costs and QALYs for exemplar stratified-NBSP.

Introduction

National breast screening programmes (NBSPs) have emerged as important public health interventions that aim to reduce deaths from breast cancer through early detection [1]. NBSPs in different jurisdictions differ in terms of the age at which screening is first offered to women in the population (start of NBSP), the interval between screens (screening interval) and the age at which screening is stopped. In the UK, the current NBSP is targeted at women within the first three years of their 50th birthday until the age of 70 years with a three-yearly screening interval [2]. In some areas of the UK, the age range has been extended to women aged 47 to 49 years and 71 to 73 years as part of an age extension trial [3]. The current UK-NBSP is a standard programme with the same screening modality (mammography) offered at the same screening interval to all women regardless of their risk of developing breast cancer.

A new concept called 'stratified screening' also known as personalised screening, is being considered to replace the existing standard, or 'one size fits all' UK-NBSP, with the aim of improving the predictive value of cancer detection and, therefore, the relative cost-effectiveness of the programme[4]. Risks of breast cancer may vary across a wide range due to familial risk, mammographic density and modifiable risk factors. The potential for improved clinical and relative cost-effectiveness is achieved by modifying the screening protocol depending on an individual's characteristics such as breast cancer risk factors or the performance of the screening modality for that individual. The introduction, or any modification to, a NBSP has an opportunity cost. It is therefore important for decision makers deciding how to allocate finite budgets for screening programmes to understand the added value of any additions to or changes to a NBSP.

A substantial, but heterogeneous, economic evidence base has been developed to quantify the potential added value of NBSP. A systematic review, conducted in 2014, identified 71 economic evaluations of relevance to breast screening in a general population of women, of these, 52 were model-based evaluations [5]. There were three studies identified that conducted model-based analyses of a stratified screening strategy. Two of these studies were based in the USA [6,7] with no relevance to healthcare systems outside that setting. One study was UK-based [8] but provided no detail on the study perspective, time horizon, nature and source of model inputs or method of analysis, which meant it is not possible to critique the relevance and quality of the results. Given the lack of an existing evidence base, it was timely to design an early model-based cost effectiveness analysis (CEA) to identify the

potential impact of introducing stratified-NBSP in the UK-setting and key drivers of the relative cost-effectiveness of different types of stratified-NBSPs.

Method

An early model-based CEA was developed to address the key criteria described in Table 1 and reported in line with published criteria [9]. The concept of an early model-based economic evaluation is used in keeping with the definition offered by Annemans et al [10]. Using an early model-based economic evaluation is in keeping with the recommendation by Sculpher et al [11] to use an iterative approach to developing economic evidence to inform the introduction of new healthcare interventions.

<Table 1 here>

Interventions

Four potential approaches (hereafter (called: risk-1; risk-2; masking; masking and risk-1) to stratified-NBSP (see Table 1) were developed as part of an European-wide collaborative project called ASSURE (Adapting Breast Cancer Screening Strategy Using Personalised Risk Estimation) [4].

Comparators

The identified relevant comparator was the current UK-NBSP (see Table 1). No screening was also identified as a comparator of interest. A pragmatic approach was taken to define 'no screening' (see Table 1).

Model conceptualisation and structure

A systematic review of economic evaluations of breast screening programmes identified no relevant existing models that could be used without extensive modification [5]. A *de novo* model structure was conceptualised, in line with published recommendations [12] and developed with input from key clinical members in the ASSURE team (n=5) and external experts (n=15). The conceptualisation process identified that the model required three components to represent: the stratification approach; breast cancer natural history with screening; and the diagnosis and treatment process following a cancer detected by screening. A discrete event simulation model was used to represent these three components. Supplementary Appendix 1 shows the model structures and descriptions in detail. The model codes, created in R statistical package, are available on request.

Model input parameters

The input parameters, with key assumptions, are now described for each of the three model components together with the values used for resource use costs and outcomes, quantified using survival and quality-adjusted life years (QALYs).

The stratification process

Performance input parameters were required for each screening modality: mammography; mammography adjusted for masking; ultrasonography and magnetic resonance imaging.

Mammography

The sensitivity of mammography was defined as the conditional probability of a tumour being detected at a mammography event given the size of tumour. This model took account of latent cancers that exist at a screening round, which were not detected, and subsequently do not present in the following interval. To obtain an estimate of screening sensitivity consistent with the presence of latent cancers in the model the screening sensitivity as defined in Weedon-Fekjaer et al [13] was used. Screening sensitivity was estimated jointly with the natural history parameters to be consistent with the presence of latent cancers that were simulated in this model. Sensitivity of mammography conditional on tumour size was parameterised as shown in Equation 1:

$$S(X) = \frac{exp\left(\frac{X-\beta_2}{\beta_1}\right)}{1+exp\left(\frac{X-\beta_2}{\beta_1}\right)} \qquad \text{Equation 1}$$

Table 2 reports the definitions for the parameters used in Equations 1 to 12.
<Table 2>

Mammography and adjustment for masking

Masking was defined as the case in which a cancer was present but not detected at screening due to the view of the cancer being obscured in the images by other tissues [14]. In mammography, masking was expected to occur more frequently when there was high breast density or if particular textural patterns of the breast tissue were present. To quantify masking due to breast density it was necessary to rely on a comparison of screen-detected and interval breast cancer rates within different density groups. From such a comparison it was possible to estimate the sensitivity of screening mammography for each group by the method of counting the screen-detected cancers as true positives and the interval cancers as false negatives.

To calculate the Volpara Density Group (VDG) specific sensitivity (Sen_{VDG}; see Equation 2 and Table 2) of mammography for a tumour of a given size, the ratio of the odds of a true positive result for that VDG compared with the population average odds (OR_{VDG}; see Equation 3 and Table 2) was combined with the odds of true positive result given tumour size alone. The resultant value for odds was then converted back to a probability to give VDG-specific and tumour size-specific sensitivity. For simplicity, it was assumed that the relative sensitivities (i.e. odds ratios) between VDGs were equal across all tumour sizes.

$$OR_{VDG} = \frac{Sen_{VDG}/(1-Sen_{VDG})}{Sen_{average}/(1-Sen_{average})} \qquad \text{Equation 2}$$

$$Sen_{X,VDG} = \frac{Sen_X/(1-Sen_x)\times OR_{VDG}}{1+(Sen_X/(1-Sen_x)\times OR_{VDG})} \qquad \text{Equation 3}$$

Mammography recall rate (true positives and false positives)

The rate of recalls that result in biopsy (true positives) was taken from a previous economic evaluation [15]. The recall rate, for women in whom no cancer is present (false positives), was calculated by identifying the overall recall rate for the UK-NBSP from published programme statistics 2011-2012 [16]. Around 20% of recalls were cited to be true positives, which indicated the estimated recall rate, excluding true positives, was 3.2%.

Ultrasonography (US) and Magnetic Resonance Imaging (MRI)

Two supplemental screening modalities were relevant. Ultrasonography (US) supplemental screening, delivered either using hand-hand equipment (HHUS) or automated equipment (ABUS) was proposed for women with high breast density (VDG3 and VDG4). For women at high risk that also have high breast density Magnetic Resonance Imaging (MRI) was used as a supplemental screening technology.

It was necessary to assume that the only available published estimates of supplemental US and MRI screening sensitivity and specificity in this group were approximately equal to those for the relevant population (mammogram negative women of screening age).The estimate of US screening performance was taken from a published systematic review and meta-analysis [17]. This review only included studies in the 'high risk' population but was the only available source that provided a quantitative synthesis of sensitivity and specificity for US. For MRI, data from an ongoing trial in a high-risk population of women in this area, Vreeman et al (personal communication), was used to inform the MRI screening performance parameters in the model.

The same approach was taken to calculate the screening performance for US and MRI. Reported cancer detection rates from each source were used to calculate the odds ratio for detecting cancer with US, MRI and mammography compared with mammography alone. The estimated odds ratio was assumed to be constant across tumour size. Equation 4 (see Table 2), shows the case for MRI:

$$OR_{MRI} = \frac{c.d.r._{\cdot mammo,MRI}/(1000-c.d.r._{\cdot mammo,MRI})}{c.d.r._{\cdot mammo}/(1000-c.d.r._{\cdot mammo})} \qquad \text{Equation 4}$$

The cancer detection rate with mammography and MRI reported by Vreemann et al was 12.14 per 1000, while the cancer detection rate for mammography alone in this group was 4.2 per 1000 [17]. The estimated odds ratio was 2.91, which was then applied to the tumour size and breast density specific odds of a cancer being detected with mammography alone. These odds can then be converted back to probabilities for use in the simulation of individual screening events using the formula in equation 5 (see Table 2):

$$Sen_{X,VDG,MRI} = \frac{Sen_{X,VDG}/(1-Sen_{x,VDG})\times OR_{MRI}}{1+(Sen_{X,VDG}/(1-Sen_{x,VDG})\times OR_{MRI})} \qquad \text{Equation 5}$$

US and MRI recall rate

The recall rate for US was 98 per 1000 exams and for MRI it was 41 per 1000 exams [17]. It was assumed that the biopsy rate for recalls is the same as the current-NBSP, which was informed by the opinion of three experts (radiologists) in the ASSURE project [4].

Breast cancer natural history with screening

Breast cancer natural history was represented using a continuous time and tumour size growth model to allow variation in growth rates. The natural history of breast cancer was defined by estimating the incidence of breast cancer with screening and the growth of tumours once detected.

Breast cancer incidence

The occurrence of breast cancer for an individual was assumed equal to the life-time risk score of that individual, estimated using the Tyrer-Cuzick algorithm [18]. This assumption implies that the risk model used in the stratification process is perfectly calibrated to the population. The age of breast cancer incidence (malignant neoplasm of breast (ICD-9: C50) and carcinoma in-situ of breast (ICD-9: D05), conditional on life-time occurrence, was then

estimated for each individual based on Office of National Statistics (ONS) cancer registry data [19](see Supplementary Appendix 2).

Breast cancer growth

A continuous time model was used to estimate the growth of tumours of the breast (see Supplementary Appendix 3). Four candidate growth models ([13,20–22]) were identified from a systematic review of economic evaluations of NBSP [5] . Each identified growth model used a unique combination of parameters, which meant a formal quantitative synthesis was not appropriate, and the natural history model published by Weedon-Fekjaer et al (2008) [13] was judged to be the best available because of: the use of a continuous growth model; the high quality of the reporting and the relatively close match in location (Europe) and time period to the current UK setting. The natural history model parameterisation was described by two steps. The parameter estimates are listed in Table 3.

<Table 3 here>

Step 1. Equations 6 (see Table 2) and Equation 7 (see Table 2) show the logistic tumour growth function (Using tumour volume V mm^3, diameter s mm, time in years t and growth rate k, and assuming a spherical shape as is in Weedon-Fekjaer et al):

$$V(t) = \frac{V_{max}}{\left[1+\left(\left(\frac{V_{max}}{V_{cell}}\right)^{0.25}-1\right)e^{-0.25kt}\right]} \qquad \text{equation 6}$$

$$s(t) = 2\left(V(t)/\left(\frac{4}{3}\pi\right)\right)^{\frac{1}{3}} \qquad \text{equation 7}$$

Step 2. Equations 8 (see Table 2) and Equation 9 (see Table 2) show the extension to individual growth rates (mixed model):

$$V_i(t) = \frac{V_{max}}{\left[1+\left(\left(\frac{V_{max}}{V_{cell}}\right)^{0.25}-1\right)e^{-0.25k_it}\right]} \qquad \text{equation 8}$$

$$s_i(t) = 2\left(V_i(t)/\left(\frac{4}{3}\pi\right)\right)^{\frac{1}{3}} \qquad \text{equation 9}$$

Diagnosis and treatment process

Following a screen detected cancer, the model captured the diagnostic and subsequent treatment process. Three types of tumours for breast cancer were reflected in the model: invasive; none or micro-invasive; advanced.

Invasive tumours

For invasive cancers, the Nottingham Prognostic Index (NPI), a commonly used and validated classification system, was used to group the diagnosed tumours into three prognostic groups [23,24]. A systematic review was used to identify reported survival for NPI-defined sub-groups [see Appendix A4]. A meta-regression analysis showed there was substantial heterogeneity between the studies that was driven by the date in which the data were collected and a trend for improved survival over time, which implied it was more appropriate to select the most recent data to inform the probability of NPI group membership conditional on invasive tumour size category and survival for women diagnosed with breast cancer (see Table 2). Allocation of invasive cancer cases to NPI categories used the probability of NPI group membership conditional on tumour size category, as reported in Kollias et al 1999 [25], which was the only study identified reporting the required cross-tabulation of size and NPI category. The required probabilities of NPI sub-group membership were calculated using the reported cross tabulation of size category by NPI category (see Supplementary Appendix 3)

None or micro-invasive tumours

Three simplifying assumptions were made to capture the impact of detecting none-invasive or micro-invasive tumours; defined as 'ductal carcinoma in-situ (DCIS)'. A vanishingly small proportion of DCIS-tumours will not be screen-detected and, therefore, it was assumed that only screen-detected cancers may be assigned to the DCIS category. The proportion of screen-detected DCIS cancers was assumed to be constant regardless of the screening interval. This assumption was supported by the proportions of DCIS in screen-detected cancers in the UK-NBSP (three-year interval; 20.3%) [26] compared with the Netherlands-NBSP (two-year interval; 20.9%) [27] being similar. Survival for DCIS diagnosed and treated patients was assumed to be the same as for the general population in line with an audit of UK screen-detected breast cancers [26]. On this basis, any screen-detected cancer was given a probability of 0.203 of being assigned to the DCIS category. DCIS cancer cases have the same all-cause survival as the general population.

Advanced tumours

A small proportion of all breast cancers will present at the advanced stage with distant metastases defined as being Stage IV in TMN classification system [28]. The probability of a

breast cancer of a given size presenting at an advanced stage was assumed not to be related to the type of screening modality or interval. The source for the probabilities of advanced breast cancer at diagnosis conditional on tumour size was taken from the NHS audit of screen-detected breast cancers (2013) (see Supplementary Appendix 3). Estimates of 10-year survival for patients with advanced breast cancer were obtained from a meta-analysis of registries in six countries [29].

Survival, invasive (non-advanced) breast cancer

For women without a diagnosis of breast cancer, survival was taken from published population life-tables [30] and the parameters of a Weibull survival distribution were estimated. Simulation of individual age of mortality was achieved by inverting the Weibull cumulative distribution function and taking a random draw from the uniform(0,1) distribution using equation 10 (see Table 2):

$$T_m = \left(\frac{\log(U)}{\lambda}\right)^{1/\upsilon} \qquad \text{equation 10}$$

The observed effect of data collection date on survival from the meta-regression [manuscript under review] meant the most appropriate estimate of survival for women with a diagnosis of breast cancer was the most up-to-date estimate (see Fong et al 2015) [31]. The parameters of four functional forms for the baseline hazard function were estimated in a regression based survival analysis: exponential, Weibull, log-normal and log-logistic. The exponential model was selected model based on the Akaike Information Criterion (AIC; a measure of model fit) and visual inspection of Cox-Snell residuals (see Supplementary Appendix 4). Estimated coefficients (Table 3) from the parametric survival model were used to simulate a survival time by inverting the survival function and use of a random number generator using equation 11 (see Table 2):

$$T_c = \frac{-log(U(0,1))}{\gamma} \qquad \text{equation 11}$$

Fong et al (2015) presented data for women aged 50-65 years, including both screen-detected and interval cancers, and it was necessary to age adjust these data for women older than 65 years of age. A further adjustment was made to account for lead time in screen-detected cancers by reversing the process of lead time bias correction as described in [32] to introduce lead time for screen-detected cancers. Mortality from breast cancer for screen-detected cancers was, therefore, calculated from the simulated time the cancer would have presented clinically rather than the time of screen detection. Standard all-cause mortality was applied in the period between screen detection and clinical presentation. This adjustment implied that an assumption was made that there was no important short-term

negative effect on mortality from treatment. It was further assumed that breast cancer did not affect survival beyond ten-years after clinical presentation (hazard rate returns to the population rate). Overall survival time post breast cancer diagnosis (T_o) was calculated using equation 12 (see Table 2):

$$if\ (T_m \leq T_d), T_o = T_m$$
$$else\ T_o = T_c + T_d \qquad \text{equation 12}$$

Quality adjusted life years

QALYs were used to capture the consequence of each screening programme. In accordance with standard practice, life-years were adjusted for average health-related quality-of-life at a given age [33]. Estimates for these age-specific average utility weights were taken from [34]. The multiplicative method was used to combine health state utility weights and age-specific average utility weights [33]. Utility weights were identified by updating a published systematic review for breast cancer health states [35]. An identical search strategy limited to the period January 2010-October 2015 yielded 11 additional studies. Consistent with the suggestions made by Peasgood et al, heterogeneity in the studies meant that meta-analysis of utility weights was inappropriate [35]. Therefore relevant utility weights were identified from studies that most closely represented the health states in the model structure. No studies were identified that defined breast cancer health states for specific NPI categories. Therefore, the selected utility weights (see table 2) were taken from Lidgren et al [36] (2007) were used for early disease and advanced (distant metastases) disease, for the first year following diagnosis and subsequent years. These selected utility weights were assumed to also account for the impact of disutility from treatment, which is in keeping with the original source for these data.

Resource Use and Costs

In accordance with the assumed healthcare system perspective, resource use and associated costs accruing to the health services were used as model input parameters (see Table 2). Initial treatment and follow-up healthcare costs were included. Costs associated with treatment for breast cancer cases of DCIS, NPI-categories 1 to 3 and advanced cancer, were taken from a published study [24] [37] [15]. These estimates from 1992 were inflated to 2015 prices using the retail price index produced by the Office of National Statistics[38]. Supplementary imaging (US and MRI) costs were taken from the NHS schedule of reference costs (2013/14) from the categories: diagnostic whole breast ultrasound (no complications), and diagnostic breast MRI (no complications). Mammography costs were sourced from [15] and reflected estimates from a screening programme. An estimate of the cost of

administering risk and breast density based stratification was made based on experience from the PROCAS study [5]. The average cost per women was estimated as £10.57 (see Supplementary appendix 1 for further details).

Data analysis

The base case analysis calculated the total costs and QALYs for a sample of 100 million women over a life time from the relevant age (in years) reflecting the start of: each of the four specified stratified breast screening programmes; current UK-NBSP; no screening. Supplementary appendix 5 shows how using a sample of 100 million women should be sufficient to be confident that the model had sufficiently converged.

Incremental analysis was performed by comparing each stratified NBSP with (i) current NBSP and (ii) no screening. In addition, a full incremental analysis was performed. All costs and QALYs were discounted at a rate of 3.5%.

One-way sensitivity analyses were used to explore the impact of selected input parameters (see Supplementary Appendix 6). In addition, NICE recommend that a relevant sensitivity analysis for interventions such as screening with long-term outcomes is to apply a 1.5% discount rate for health outcomes and a 3.5% discount rate for costs [39]. In common with previously published economic evaluations in screening, a no discounting scenario was also estimated. Probabilistic sensitivity analysis (PSA) [40] was performed to quantify the effect of the joint uncertainty (see Supplementary Appendix 7) using a generalised additive model [41].

Results

Table 4 shows the results of the base case analysis for a: risk-based stratified-NBSP (using Risk-1 or Risk-2); masking-based stratified-NBSP (Masking); risk and masking-based stratified-NBSP (Risk-1 and Masking). The Risk-1 stratified-NBSP and Risk-2 stratified-NBSP were relatively cost-effective when compared with the current UK-NBSP. The Masking stratified-NBSP does not appear to be a cost-effective alternative when compared with the current UK-NBSP. Using an alternative discounting rate of 3.5% for costs and 1.5% for benefits, resulted in relatively lower estimated ICERs for all stratified-NBSP compared with the UK-NBSP. When compared with no screening, all screening programmes may be considered cost-effective. A full incremental analysis is available in Supplementary Appendix 7. This shows that masking and risk-1 and masking was dominated by the next alternative (current-NBSP and risk-1 stratified NBSP, respectively). The ICER for the remaining comparisons were: £23,197 per QALY for the current-NBSP compared with no screening;

£16,689 per QALY for risk-1 stratified NBSP compared with masking; £26,749 for risk-2 stratified NBSP compared with masking and risk-1 stratified NBSP.

<Table 4 here>

Sensitivity Analyses

To examine the decision between using the suggested stratified-NBSP and the current UK-NBSP a cost-effectiveness acceptability curve is presented in Figure 1 using the results of the PSA. Figure 2 shows the associated cost-effectiveness acceptability frontier, that suggests the current UK-NBSP would be selected as the preferred programme with a threshold of cost per QALY gained below £20,000 per QALY gained while the Risk-2 stratified NBSP would be chosen at higher thresholds of cost per QALY gained.

<Figure 1 here>
<Figure 2 here>

One-way sensitivity analysis (see Supplementary Appendix 6) showed that the reported total costs, total QALYs and ICERs were sensitive to: natural history parameter values (α_2 and mean tumour size at clinical detection) and screening performance of mammography (β_2). ICERs for stratified programmes were moderately sensitive to the cost of stratification although costs would need to be several times the base case value for ICERs to in increase beyond a threshold of £30,000 per QALY. In all alternative programmes total costs were sensitive to the treatment cost parameters however varying these parameters did not greatly change the ICERs compared with the base case. Estimates of total QALYs were sensitive to the utility weights for cancer states, varying utility weights moderately altered the ICERs of stratified programmes compared with the NBSP. The results were relatively insensitive (within the ranges tested) to: the probability of recall; costs of MRI; the relative sensitivity of mammography by VDG group; and US/MRI additional cancer detection rate.

Discussion

This study used an early model-based CEA to generate estimates of the relative costs and consequences of four example stratified-NBSPs compared with no screening and current practice in the UK-NBSP. The Risk-1 and Risk-2 stratified-NBSP compared to the current UK-NBSP were deemed to be a cost-effective use of healthcare resources relative to a threshold range of £20,000 to 30,000 per QALY gained. The ICERs for the current UK-NBSP compared with no screening were somewhat higher than previous analyses [15,42] but were very similar to the most recently published study [43]. Results were not directly comparable

to previous model-based analyses of stratified screening [6,7] due to differences in modelling strategy and also comparators.

The Masking stratified-NBSP was relatively the less cost-effective strategy. Combining the two stratification approaches using Risk-1 and Masking simultaneously resulted in modest QALY gain when compared with either Risk 1 or Masking stratified NBSP. The modest gains from masking based strategies could be due to increased over-diagnosis overwhelming the potential QALY gains from early detection of a tumour. Over-diagnosis is a commonly cited problem with NBSP [1]. Over-diagnosis suggests that NBSPs are too effective at detecting small, and slow growing, tumours that would not affect a woman's health within her lifetime if left undetected. Follow up procedures such as biopsies and treatment for such over-diagnosed cases are expensive and may cause harm [44].

The interpretation of the cost-effectiveness results for stratified breast screening was strongly influenced by the choice of discount rate. The choice of discount rate is not a simple technical question and the preferred discounting procedure for producing cost-effectiveness results for economic evaluations in health is a contested issue [39]. Decision makers should consider which discounting scenario best reflects the values and preferences of those for whom they are making a decision,

This early economic analysis was based on the best available data sourced from a combination of rapid reviews, systematic reviews and analysis of data from two key published prospective studies [31,45]. Key data gaps were: the relative sensitivity of mammography by density given the tumour size; the detection rate of supplemental ultrasound; recall rate and biopsy rate. Most importantly, the lack of randomised trials, or sufficiently long robust observational studies, meant that there were no direct estimates of the effect of supplemental screening modalities on mortality or other long-term outcomes. Robust, up-to-date, data on the cost of treating women with breast cancer were not available. This meant it was necessary to rely on estimates from a now dated study for the cost of treatment stratified by a prognostic indicator [37]. In addition, on the advice of clinical experts the implications of screening on use of different targeted treatment options based tumour HER2 or ER receptor status was not included in this model. These important uncertainties, due to the lack of robust data for several key parameters suggest that the results of this model-based CEA should be treated as indicative. The focus should be on the model structure itself and on the identified key drivers of relative cost-effectiveness. The most important drivers of cost-effectiveness after the discount rate were the natural history parameters, cost of stratification and mammographic sensitivity parameters and future research should be directed at improving the robustness of these data. Some one-way

sensitivity analysis results may appear inconsistent (US and MRI cancer detection rates) and this may be due to Monte Carlo error for alternatives where the differences between strategies were small or the result of non-linearity in the model.

Decision makers using the results of this DES model-based CEA must recognise the inherent limitations of mathematical models of disease natural history and screening that may introduce structural uncertainty. Using a DES, was in line with published models in cancer screening [46]. DES allowed the influence of individual patient-characteristics to be captured; the flexibility for cancer growth to be modelled as a continuous process; the use of prognostic categories to group treatment options. It may be that modelling choices, such how cancer growth rates can vary between individuals, were influential in driving the relative cost-effectiveness of a particular NBSP. No formal external validation or calibration of this early decision analytic model was conducted. External validation against more extensive clinical trial or observational data should be a goal of any future investigation of the cost-effectiveness of stratified-NBSP.

A key important assumption, in the absence of data to prove otherwise, was that the risk model used in the stratification process was perfectly calibrated to the population. This 'structural' uncertainty is not reflected results of PSA therefore users must exercise judgement when interpreting the results. A further limitation to be aware of is that the use of regression models within a PSA is a new and developing methodology therefore these results should perhaps be treated with some caution. Structural uncertainty may be best addressed by planning external validation studies in future research relating to all aspects of the economic model, including, the risk models used in stratification and the natural history models of breast cancer. External validation studies of the risk model to be used in a stratified-NBSP are essential if there is reason to believe calibration may be poor, which also requires consensus to be reached on the appropriate risk categories to use in practice. Previous experience in a research context suggests that embedding stratification in the existing NBSP is feasible [47] but no data exist on the effects of stratification on screening uptake and this is an important topic for further research.

Conclusions

This early model-based CEA presents indicative results that suggest a risk stratified-NBSP is potentially a cost-effective use of healthcare resources when compared to the current UK-NBSP. The proposed model structure will be a key resource as more data become available to support the introduction of stratified-NBSP such as the sensitivity and effectiveness of the

new screening modalities, the effect of risk communication strategies on NBSP uptake and the cost of newer treatments for breast cancer. The choice of discount rate will be crucial in interpreting the results. A pre-specified external validation analysis should be conducted alongside any more definitive economic evaluation.

References

[1] Marmot M, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M. The Benefits and Harms of Breast Cancer Screening: An Independent Review. London, UK: Cancer Research UK and the Department of Health (England); 2012.

[2] Breast screening: programme overview 2015. https://www.gov.uk/guidance/breast-screening-programme-overview#screeningage (accessed May 30, 2016).

[3] Breast screening: age extension trial 2015. https://www.gov.uk/government/publications/nhs-breast-screening-age-extension-trial (accessed May 30, 2016).

[4] ASSURE: Personalised Breast Screening 2014. http://www.assure-project.eu/ (accessed May 30, 2016).

[5] Evans G, Stavrinos P, Dawe S, Harvie M, Wilson M, Maxwell A, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and Family History Clinic: a dual cohort study. 2015.

[6] Tosteson ANA, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-Effectiveness of Digital Mammography Breast Cancer Screening. Ann Intern Med 2008;148:1. doi:10.7326/0003-4819-148-1-200801010-00002.

[7] Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing Mammography by Breast Density and Other Risk Factors for Breast Cancer : Analysis of Health Benefits and Cost-Effectiveness. Ann Intern Med 2011;155:10–21.

[8] Van Dyck W, Gassull D, Vértes G, Jain P, Palaniappan M, Schulthess D, et al. Unlocking the value of personalised healthcare in Europe—breast cancer stratification. Heal Policy Technol 2012;1:63–8. doi:10.1016/j.hlpt.2012.04.006.

[9] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMC Med 2013;11:80. doi:10.1186/1741-7015-11-80.

[10] Annemans L, Genesté B, Jolain B. Early Modelling for Assessing Health and Economic Outcomes of Drug Therapy. Value Heal 2000;3:427–34. doi:10.1046/j.1524-4733.2000.36007.x.

[11] Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy 1997;2:26–30. doi:10.1177/135581969700200107.

[12] Roberts M, Russell LB, Paltiel a D, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Med Decis Mak 2012;32:678–89. doi:10.1177/0272989X12454941.

[13] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Res 2008;10:R41. doi:10.1186/bcr2092.

[14] Harvey JA, Bovbjerg VE. Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk1. Radiology 2004;230:29–41. doi:10.1148/radiol.2301020870.

[15] Madan J, Rawdin A, Stevenson M, Tappenden P. A rapid-response economic evaluation of the UK NHS Cancer Reform Strategy breast cancer screening program extension via a plausible bounds approach. Value Health 2010;13:215–21. doi:10.1111/j.1524-4733.2009.00667.x.

[16] Patnick J. NHS Breast Screening Programme Annual Review 2012. 2012.

[17] Tice J, Ollendorf D, Lee J, Pearson S. The Comparative Clinical Effectiveness and Value of Supplemental Screening Tests Following Negative Mammography in Women with Dense Breast Tissue. 2013.

[18] Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 2004;23:1111–30. doi:10.1002/sim.1668.

[19] Office of National Statistics. Cancer Incidence and Mortality in the United Kingdom - 2008-10. 2012.

[20] Baker RD. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Manag Sci 1998;1:103–13.

[21] Van Oortmarssen GJ, Habbema JDF, Van Der Maas PJ, De Koning HJ, Collette HJ a., Verbeek ALM, et al. A model for breast cancer screening. Cancer 1990;66:1601–12. doi:10.1002/1097-0142(19901001)66:7<1601::AID-CNCR2820660727>3.0.CO;2-O.

[22] Tan KHX, Simonella L, Wee HL, Roellin A, Lim Y-W, Lim W-Y, et al. Quantifying the natural history of breast cancer. Br J Cancer 2013;109:2035–43. doi:10.1038/bjc.2013.471.

[23] Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361–6.

[24] Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat 1992;22:207–19. doi:10.1007/BF01840834.

[25] Kollias J, Murphy CA, Ellis IO, Elston CW, Robertson JFR, Blamey RW. The prognosis of small primary breast cancers. Eur J Cancer 1999;35:908–12.

[26] [UK NHSBSP and ABS Screening Audit Group]. NHS breast screening programme and association of breast surgery; An audit of screen detected breast cancers for the year of screening April 2013 to March 2014. 2015.

[27] Bluekens AMJ, Holland R, Karssemeijer N, Broeders MJM, den Heeten GJ. Comparison of Digital Screening Mammography and Screen-Film Mammography in the Early Detection of Clinically Relevant Cancers: A Multicenter Study. Radiology 2012;265:707–14. doi:10.1148/radiol.12111461.

[28] Sobin LH, Gospodarowicz MK (Mary K., Wittekind C (Christian), International Union against Cancer. TNM classification of malignant tumours. Wiley-Blackwell; 2009.

[29] Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007. Thorax 2013;68:551–64. doi:10.1136/thoraxjnl-2012-202297.

[30] Office of National Statistics. United Kingdom, National Life Tables, 1980-82 to 2011-13. 2014.

[31] Fong Y, Evans J, Brook D, Kenkre J, Jarvis P, Gower-Thomas K. The Nottingham Prognostic Index: five- and ten-year data for all-cause survival within a screened population. Ann R Coll Surg Engl 2015;97:137–9. doi:10.1308/003588414X14055925060514.

[32] Duffy SW, Nagtegaal ID, Wallis M, Cafferty FH, Houssami N, Warwick J, et al. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. Am J Epidemiol 2008;168:98–104. doi:10.1093/aje/kwn120.

[33] Ara R, Wailoo A. NICE DSU Technical Support Document 12: the use of health state utility values in decision models. … , UK Sch Heal Relat Res … 2011:1–30.

[34] Ara R, Brazier JE. Using Health State Utility Values from the General Population to Approximate Baselines in Decision Analytic Models when Condition-Specific Data are Not Available. Value Heal 2011;14:539–45. doi:10.1016/j.jval.2010.10.029.

[35] Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev Pharmacoecon Outcomes Res 2010;10:553–66. doi:10.1586/erp.10.65.

[36] Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Qual Life Res 2007;16:1073–81. doi:10.1007/s11136-007-9202-8.

[37] Johnston K. Modelling the future costs of breast screening. Eur J Cancer 2001;37:1752–8. doi:10.1016/S0959-8049(01)00202-7.

[38] Office of National Statistics. Consumer Price Indices - RPI indices: 1987 to 2015. 2015.

[39] Jonsson P, Robertson J. Briefing paper for methods review working party on discounting. 2013.

[40] Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339–47. doi:10.1002/hec.985.

[41] Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making 2014;34:311–26. doi:10.1177/0272989X13505910.

[42] Forrest P. Breast Cancer Screening, Report to Health Ministers of England, Wales, Scotland & Northern Ireland 1986.

[43] Pharoah PDP, Sewell B, Fitzsimmons D, Bennett HS. Cost effectiveness of the NHS breast screening programme : life table model. Br Med J 2013;2618. doi:10.1136/bmj.f2618.

[44] Alvarado M, Ozanne E, Esserman L. Overdiagnosis and overtreatment of breast cancer. Am Soc Clin … 2011.

[45] Evans DGR, Donnelly LS, Harkness EF, Astley SM, Stavrinos P, Dawe S, et al. Breast cancer risk feedback to women in the UK NHS breast screening population. Br J Cancer 2016;114:1045–52. doi:10.1038/bjc.2016.56.

[46] Pilgrim H, Tappenden P, Chilcott J, Bending M, Trueman P, Shorthouse A, et al. The costs and benefits of bowel cancer service developments using discrete event simulation. J Oper Res Soc 2008;60:1305–14. doi:10.1057/jors.2008.109.

[47] Evans D, Astley S, Stavrinos P, Harkness E, Donnelly L, Dawe S, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and Family History Clinic: A dual cohort study. 2016.

List of Figures

Figure 1: Cost effectiveness acceptability curve for all stratified-NBSPs and comparators

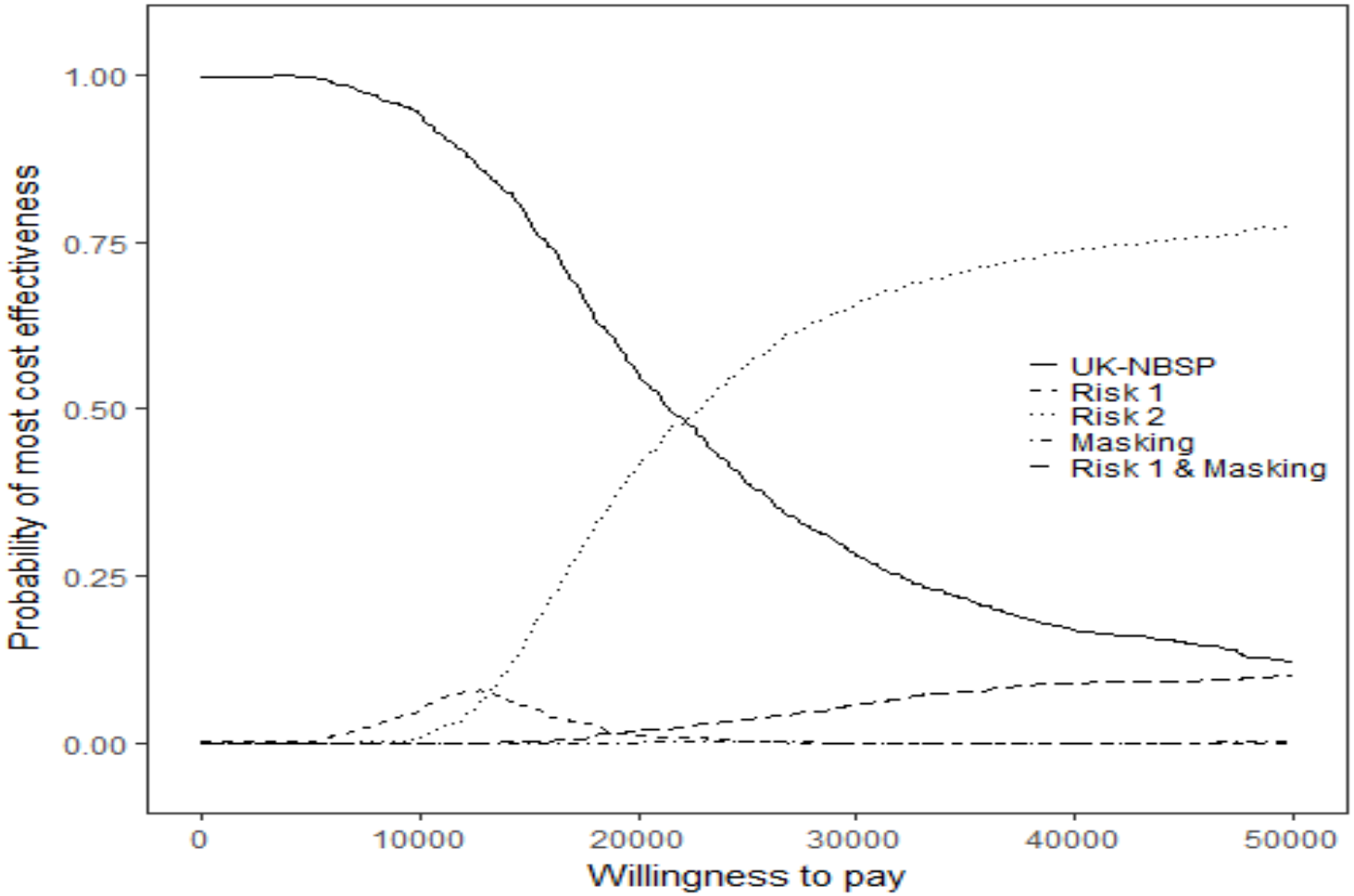
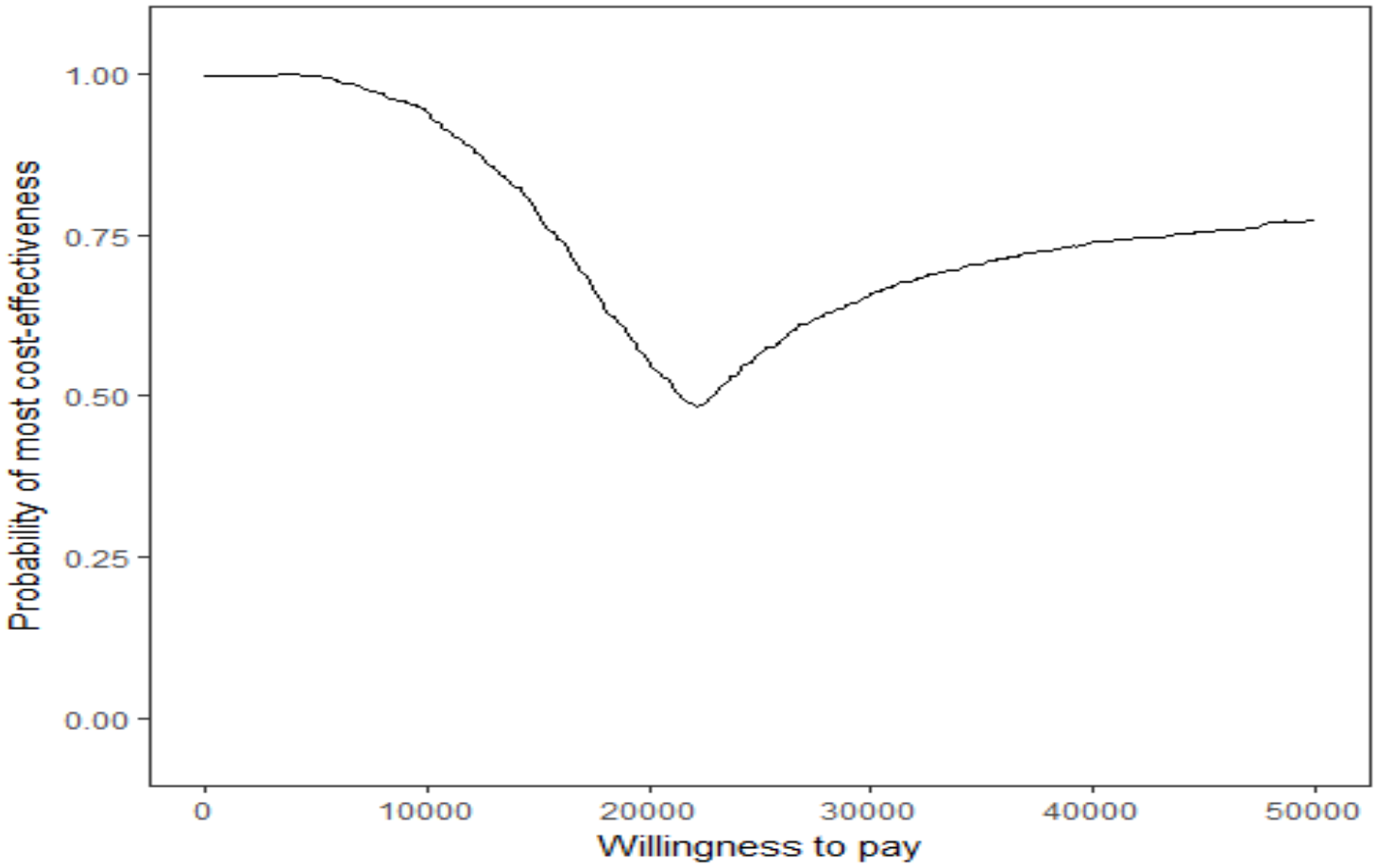


Figure 2: Cost effectiveness acceptability frontier for all stratified-NBSPs and comparators



List of Tables

Table 1: Key design criteria

Decision problem	What are the key drivers of the incremental costs and benefits of example stratified breast screening programmes compared with the current national breast screening programme?
Interventions	(i)Risk-1: a risk-based stratification defined by the risk-algorithm used in a published study [5] enhanced with density and texture measures following the method of Brentnall et al [44]. Three strata (with associated screening intervals) were defined by ten-year risks of breast cancer of (i) <3.5% (three-yearly); (ii) 3.5 to 8% (two-yearly); (iii) >8% (annual). (ii) Risk-2: a risk-based stratification defined by the same algorithm as risk-1 but with strata defined by dividing the population into thirds based on ten-year risk (tertiles): (i) the lowest risk tertile (three-yearly); (ii) the middle tertile (two-yearly); (iii) the highest risk tertile (annual). (iii) Masking (*covering up of tumours in mammograms by dense breast tissue)*: current screening approach with supplemental ultrasound offered to women with high breast density, defined using Volpara Density Grades (VDG3 and VDG4)[45]. High risk was defined as greater than an 8% ten-year risk of breast cancer [46]. Women with both high breast density and high risk of breast cancer were offered supplemental MRI instead of ultrasound. (iv) Risk-1 with masking: the risk-1 stratification approach together with the strategy described in the masking approach.
Comparators	(i)Current national breast screening programme (UK-NBSP): Women between 50 and 70 years with screening every three-years using mammography (ii) No screening: no use of mammography in the population for screening purposes. All cancers would present with clinical signs or symptoms
Model type	Discrete event simulation programmed in R
Population	Women eligible for a national breast screening programme
Setting and perspective	National healthcare service Costs to individual women were excluded from the analysis
Time Horizon	lifetime
Costs	National currency (£) at 2014 prices
Benefits	Life-years and quality adjusted life years
Discounting	3.5% for both costs and benefits (base case) 3.5% for costs and 1.5% for benefits (sensitivity analysis)
Cost-effectiveness threshold	NICE UK-recommended threshold of £20,000 per QALY gained

Table 2: List of parameters and definitions for equations

Equation	Parameter	Definition
Equation 1	$S(X)$,	sensitivity of mammography to detect a tumour of size X (maximum diameter in mm)
	β_1	sensitivity of mammography; β_1 determined how rapidly sensitivity changes with tumour size to approach the asymptotes of 0 and 1.
	β_2	sensitivity of mammography; β_2 places the location of the sensitivity curve in relation to tumour size, where $X - \beta_2 = 0$ sensitivity is equal to 0.5.
Equation 2	$Sen_{average}$	sensitivity of mammography without density information
	Sen_X	sensitivity of mammography given a tumour size X
	$Sen_{X,VDG}$	size and Volpara Density Group (VDG) specific sensitivity of mammography
Equation 3	$Sen_{average}$	sensitivity of mammography without density information
	Sen_X	sensitivity of mammography given a tumour size X
	$Sen_{X,VDG}$	size and Volpara Density Group (VDG) specific sensitivity of mammography
Equation 4	OR_{MRI}	the odds ratio for detecting cancer with MRI and mammography compared with mammography alone
	$c.d.r._{\cdot mammo,MRI}$	the cancer detection rate for the combined methods
	$c.d.r._{\cdot mammo}$	the cancer detection rate for mammography alone
Equation 5	$Sen_{X,VDG,MRI}$	the sensitivity of screening with mammography and MRI for a tumour of size X in a women classified as in a given Volpara Density Group (VDG)
	$Sen_{X,VDG}$	the sensitivity for the same tumour for mammography alone
Equation 6	V_{max}	the assumed maximum tumour volume, equal to a sphere of 128mm diameter
	V_{cell}	the assumed initial volume of an incident cancer, equal to a sphere of 0.025mm diameter
Equation 7	V_{max}	the assumed maximum tumour volume, equal to a sphere of 128mm diameter
	V_{cell}	the assumed initial volume of an incident cancer, equal to a sphere of 0.025mm diameter
Equation 8	k_i	the individual growth rate parameter and follows a lognormal distribution $\ln N(\alpha_1, \alpha_2)$. Individual growth rates are drawn from a lognormal distribution with mean α_1 and standard deviation α_2
	V_{max}	the assumed maximum tumour volume, equal to a sphere of 128mm diameter
	V_{cell}	the assumed initial volume of an incident cancer, equal to a sphere of 0.025mm diameter
Equation 10	T_m	Survival time (age)
	λ	scale parameter (=0.897)
	υ	shape parameter (=86.74).
	U	uniform(0,1) random draw
Equation 11	T_c	the survival time in years
	γ	the exponential survival function parameter, estimated in the parametric survival analysis, for a specific Nottingham Prognostic Indicator group
Equation 12	T_d	the time to simulated clinical detection

	T_m	the previously calculated all cause survival time, and T_c is the post-cancer diagnosis all-cause survival time.

Table 3: Input parameters for base case analysis

Parameter	Value	Source
Breast cancer risk factors	varied	Random sample from individual patient data in [45][47](PROCAS study)]
Summary statistics risk factors:		
	Mean (s.d.)	[45]
Age	48.93 (1.09)	[45]
10-year risk	3.04% (1.43)	[45]
Lifetime risk	13.21% (1.43)	[45]
Density (Volpara)	8.02% (5.26)	[45]
Cancer Incidence Parameters:		
Conditional on breast cancer in lifetime, probability it originates at age t	See supplementary appendix	[19]
Cancer growth parameters:		
Tumour starting size (diameter)	0.25mm	[13]
Maximum tumour size	128mm	[13]
Growth rate mean (log-normal) α_1	1.07	[13]
Growth rate standard deviation α_2	1.31	[13]
All-cause mortality:		
Weibull shape	8.97	Fit to life table for UK population [30]
Weibull scale	86.74	Fit to life table for UK population [30]
Mammography:		
Sensitivity by tumour size modelled as logistic-type function β_1 – sets increase with size β_2 – sets sensitivity relative to size	 1.47 6.51	[13]
Maximum sensitivity	0.95%	[13]
Sensitivity by VDG, used to calculate relative sensitivity given tumour size:		
Sensitivity VDG1	85.0%	[48]
Sensitivity VDG2	77.6%	[48]
Sensitivity VDG3	69.0%	[48]
Sensitivity VDG4	58.6%	[48]
Recall rate	4.0 per 100 exams	[16]
False positive biopsy proportion	2.4%	[16]
Proportion of screen-detected cancers that are DCIS	20.3%	[26]
Clinically detected (interval cancers):		
Cancer size at clinical detection – mean	6.5 doublings (22.62mm)	[20]
Cancer size at clinical detection standard deviation -	0.535 doublings	[20]
Survival post breast cancer diagnosis:		
γ NPI 1	-5.413	[31]
γ NPI 2	-4.023	[31]
γ NPI 3	-2.465	[31]
γ Advanced cancer, age <50	-0.527	[29]
γ Advanced cancer, age 50-69	-0.537	[29]
γ Advanced cancer, age ≥70	-0.849	[29]
US cancer detection:		

VDG3/4 incremental cancers detected with supplemental US	3 per 1000 exams	[17]
False positive (recall) rate, US	98 per 1000 exams	[17]
Biopsy rate, US	2.4%	Assumed same as mammography
Proportion cancers detected by supplemental US that are DCIS	21%	Assumed same as mammography
MRI cancer detection:		
VDG3/4 incremental cancers detected with supplemental US	5 per 1000 exams	Vreemann et al (personal communication)
False positive (recall) rate, MRI	41.15 per 1000 exams	Vreemann et al (personal communication)
Biopsy rate, MRI	3.03%	Vreemann et al (personal communication)
Proportion cancers detected by supplemental MRI that are DCIS	14.3%	Vreemann et al (personal communication)
Costs:		
Mammography	£54	[15]
Follow-up (mean)	£95	[15]
Biopsy (mean)	£160	[49]
NPI 1 treatment (mean)	£11630	[15]
NPI 2 treatment (mean)	£12978	[15]
NPI 3 treatment (mean)	£15405	[15]
Advanced cancer (mean)	£23449	[13]
Screening ABUS	£80	Expert opinion
Screening HHUS	£80	Expert opinion
Screening MRI	£220	[49]
Stratification process	£10.57	[5] and expert opinion
Utility:		
Early breast cancer – first year	0.696	[36]
Early breast cancer – subsequent years	0.779	[36]
Advanced breast cancer – first year	0.685	[36]
Advanced breast cancer – subsequent years	0.685	[36]

Table 4: Base case deterministic analyses of example stratified-NBSP

Screening Programme	QALYs (3.5% discount rate)	Cost (£; 2015; 3.5% discount rate)	Incremental cost-effectiveness ratio (ICER)			
			versus no screening (3.5% discount rate)	versus UK-NBSP (3.5% discount rate)	versus no screening (1.5% health, 3.5% costs)	versus UK-NBSP (1.5% health, 3.5% costs)
No screening[a]	17.6919	246	Not applicable	Not applicable	Not applicable	Not applicable
Current UK-NBSP	17.7095	654	£23,197	Not applicable	£11,343	Not applicable
Risk-1[b]	17.7119	694	£22,413	£16,689	£11,363	£11,565
Risk-2[c]	17.7181	858	£23,435	£23,924	£11,425	£11,592
Masking[e]	17.7102	809	£30,772	£212,947	£15,065	£105,412
Risk-1 and Masking[f]	17.7124	870	£30,532	£75,254	£14,707	£33,199

[a] No mammography used in the population for screening purposes and all cancers would present with clinical signs or symptoms.

[b] Risk-based stratification with three strata as defined by a published risk-algorithm [16] for 10-year risks of breast cancer and associated screening intervals: <3.5% with 3-yearly screening interval; 3.5 to 8% with 2-yearly screening interval; >8% with annual screening.

[c] Risk-based stratification with three strata defined by dividing the population into thirds based on risk (tertiles): lowest risk tertile with 3-yearly screening; middle tertile with 2-yearly screening interval; highest risk tertile with annual screening.

[e] Current UK-NBSP with supplemental ultrasound offered to women with high breast density. Women with both high breast density and high risk of breast cancer offered supplemental MRI instead of ultrasound. High breast density was defined using Volpara Density Grades (VDG3 and VDG4) and high risk was defined > 8% 10-year risk of breast cancer based on NICE definition of high risk (30% lifetime risk ≈ 8% 10-year risk) [46].

[f] Risk-based stratification [with three strata as defined by a published risk-algorithm [16] for 10-year risks of breast cancer and associated screening intervals: <3.5% with 3-yearly screening interval; 3.5 to 8% with 2-yearly screening interval; >8% with annual screening] AND current UK-NBSP with supplemental ultrasound offered to women with high breast density. Women with both high breast density and high risk of breast cancer were offered supplemental MRI instead of ultrasound. High breast density was defined using Volpara Density Grades (VDG3 and VDG4) and high risk was defined as greater than 8% 10-year risk of breast cancer.

Evaluation of a stratified-NBSP in the UK

Online Supplementary Appendices

Appendix 1

Model structure and description

This document describes the three components of the total model structure: the stratification process (Figure A1.1); cancer natural history and screening (Figure A1.2); diagnosis and treatment (Figure A1.3).

Stratification process model
The stratification process was assumed to operate in two stages. Firstly, upon attending her first screening appointment, each woman's risk will be assessed using a risk estimation model. The women is then informed of this risk and assigned to a risk category. Secondly, the mammographic breast density will be calculated by an automated measure immediately following the mammogram examination. The woman is informed of this result, and if density is above a pre-defined threshold then she will be offered a supplemental screening modality immediately. The interval until the woman is next invited for screening was determined by the risk category to which she was assigned.

Stratification process costs

The stratification process was assumed to cost £10.57 in the base case analysis. This was based on an assumption about the resources used in the stratification process, which was informed by the process used in the PROCAS study [1], and included those listed in the table A1.1

Table A1.1: Resource use and cost of stratification process

Resource	Quantity	Unit cost; £/2015 (source)	Component cost
Nurse visit	5 minutes	£45 per hour (PSSRU unit costs 2015)	£3.75
Data entry, grade 4 admin	3 sets of data per 5 minute	£28.80 per hour (PSSRU unit costs 2015)	£0.80
Consultant led risk counselling	20 minutes (10% of attendees	£50.20 (NHS reference costs 2014/15)	£5.02
Postage and stantionary	1 letter and additional page in invitation letter	£1 (expert opinion)	£1
Total cost per woman			£10.57

Cancer natural history and screening model
One of the most important choices in structuring the model was how to describe the natural history of breast cancer. This was particularly important because the ability of screening to affect health will depend not only on the performance of the test but also on the prevalence of early stage cancers and their characteristics. Natural history models for breast cancer have been developed by other researchers in several published analyses [2–6]. These models, along with published critiques, informed the model created for this evaluation.

The identified previous economic and epidemiological analyses of breast cancer have proposed simplified models of cancer growth. Two general approaches have been taken in the published literature:

1. State-transition model (also called Markov models): In which, cancers are assumed to grow by advancing through discrete size categories (e.g. 5-10mm) with fixed probabilities of a transition within a discrete time period (e.g. one-year).

2. Continuous growth model: A growth function is specified that is continuous in time. Tumour growth from starting size can be calculated for any future point in time.

The advantages of state-transition models are that they are potentially easier to use and make calculating model parameters more feasible when available data are in discrete categories (e.g. tumour size data, survival time conditional on tumour size etc). A major disadvantage is the greater approximation because of the introduction of discrete categories for tumour size and time. Another disadvantage is the complexity required to introduce heterogeneity in growth rates. This is because the probability of transition from one size category to the next will then depend on previous transition probabilities. This violates the Markovian ('memoryless') property of the health states that makes state-transition models mathematically tractable. The solution requires creating many more health states creating an extremely complex and unwieldy model.

Continuous growth models have the advantage that no approximations are introduced by forcing tumour size into categories with transitions at fixed intervals of time. Usually, fewer parameters will be needed in the model compared with those included in the state transition model (these require transition probabilities estimated for every possible transition). A disadvantage is that a functional form for cancer growth must be assumed. Additional assumptions will be required if the data available to estimate the growth function are discrete.

Variation in growth rates can be simulated in a continuous time model by sampling individual level growth rate parameters from an assumed population distribution. This is simple to achieve mathematically and computationally. The difficulty of achieving variation in individual's growth rates in a state-transition model gives a clear advantage to continuous growth models.

A more extensive discussion of alternative models of cancer growth for use in cost-effectiveness analysis is available in a paper by Karnon and colleagues [7]. The authors noted that despite the well-known theoretical merits of each modelling approach there has been no empirical investigation of impact on model outputs of such choices in typical cancer screening model scenarios.

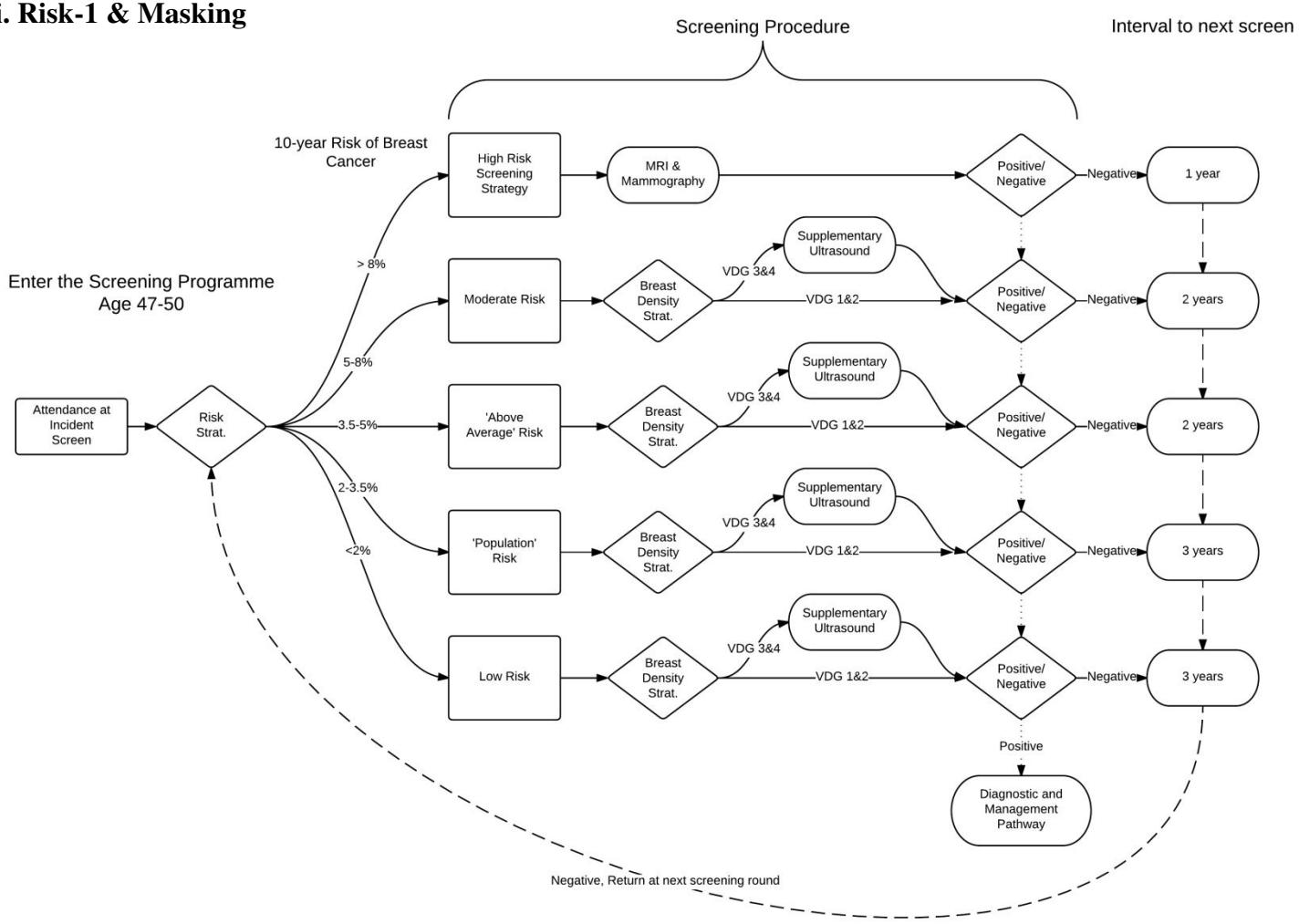
To include variation in growth rates, a continuous time and tumour size growth model was used to represent natural history. This approach was implemented using a (individual-level) discrete event simulation (DES) model.

Diagnosis, Survival and treatment

The diagnosis was linked to a process to capture prognostic information. At time of diagnosis, the size (maximum diameter measured in mm) of the cancer was defined in the model. Other prognostic factors were not set prior to diagnosis within the simulation. At a diagnosis event additional prognostic factors were simulated. The model was based around using the Nottingham Prognostic Index (NPI) [8]. NPI was selected as the basis for prognostic information modelling because of the widespread clinical and research use of this measure, including international validation [9,10]. The NPI scores three prognostic factors: tumour size (maximum diameter in millimetres); histological grade (1 to 3-point scale) and lymph node staging (1 to 3-point scale). The total scores are then commonly used to categorise individuals into prognostic groups. Three and five group classifications of prognostic groups have been commonly used. Survival data for each group can be analysed for the purpose of predicting and simulating survival. In this model made use of the three group classification for NPI.

Survival post cancer diagnosis was simulated based on NPI group. Survival 10 years post cancer diagnosis was assumed to be the same as the general population. Treatment costs were assigned as a single tariff, with the value conditional on NPI group. The costs were applied at the time of cancer diagnosis.

Figure A1.1 – Stratified Programmes
i. Risk-1 & Masking



VDG: Volpara density group

Figure A1.1 – Stratified Programmes
ii. Risk-1

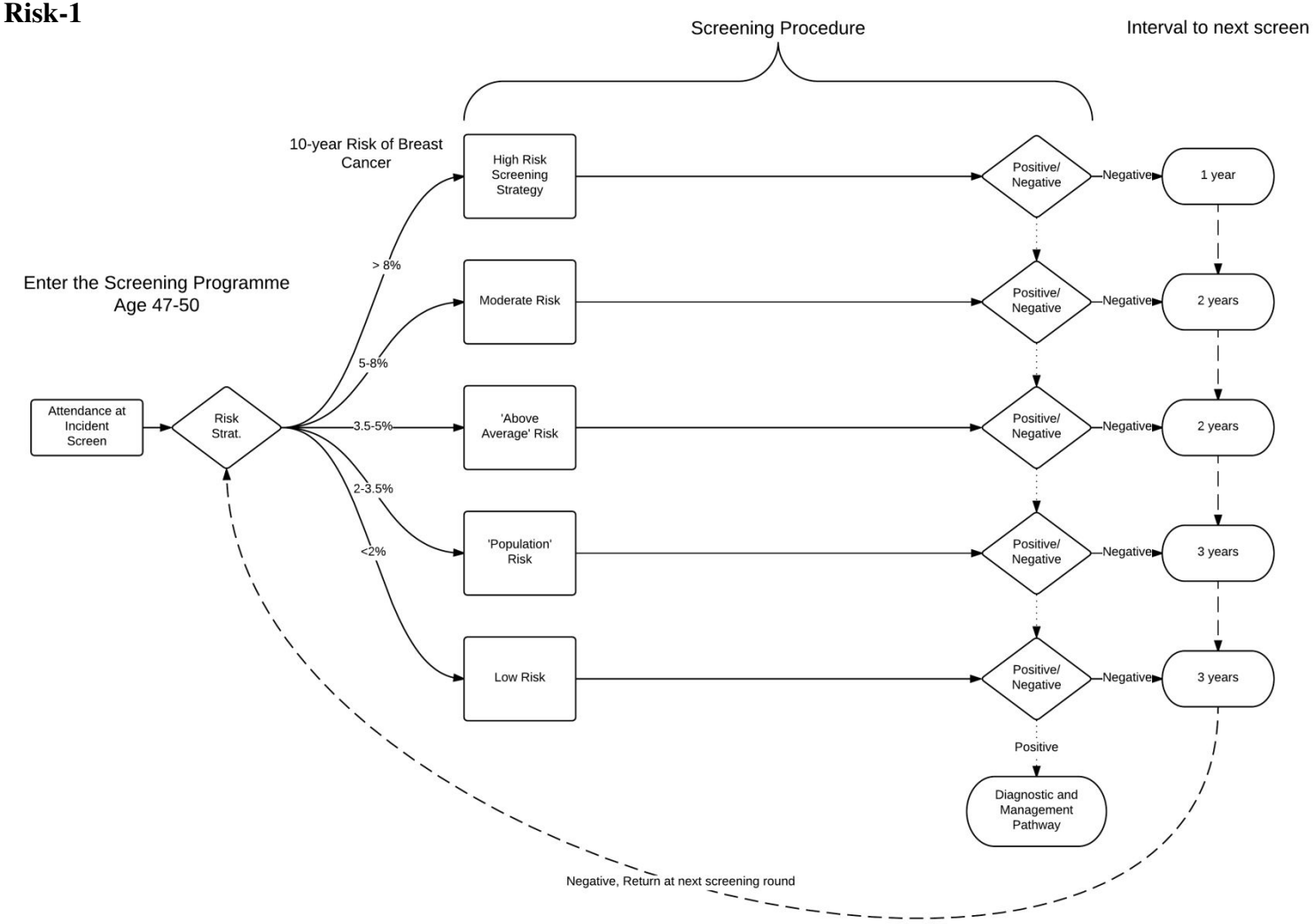


Figure A1.1 – Stratified Programmes
iii. Risk-2

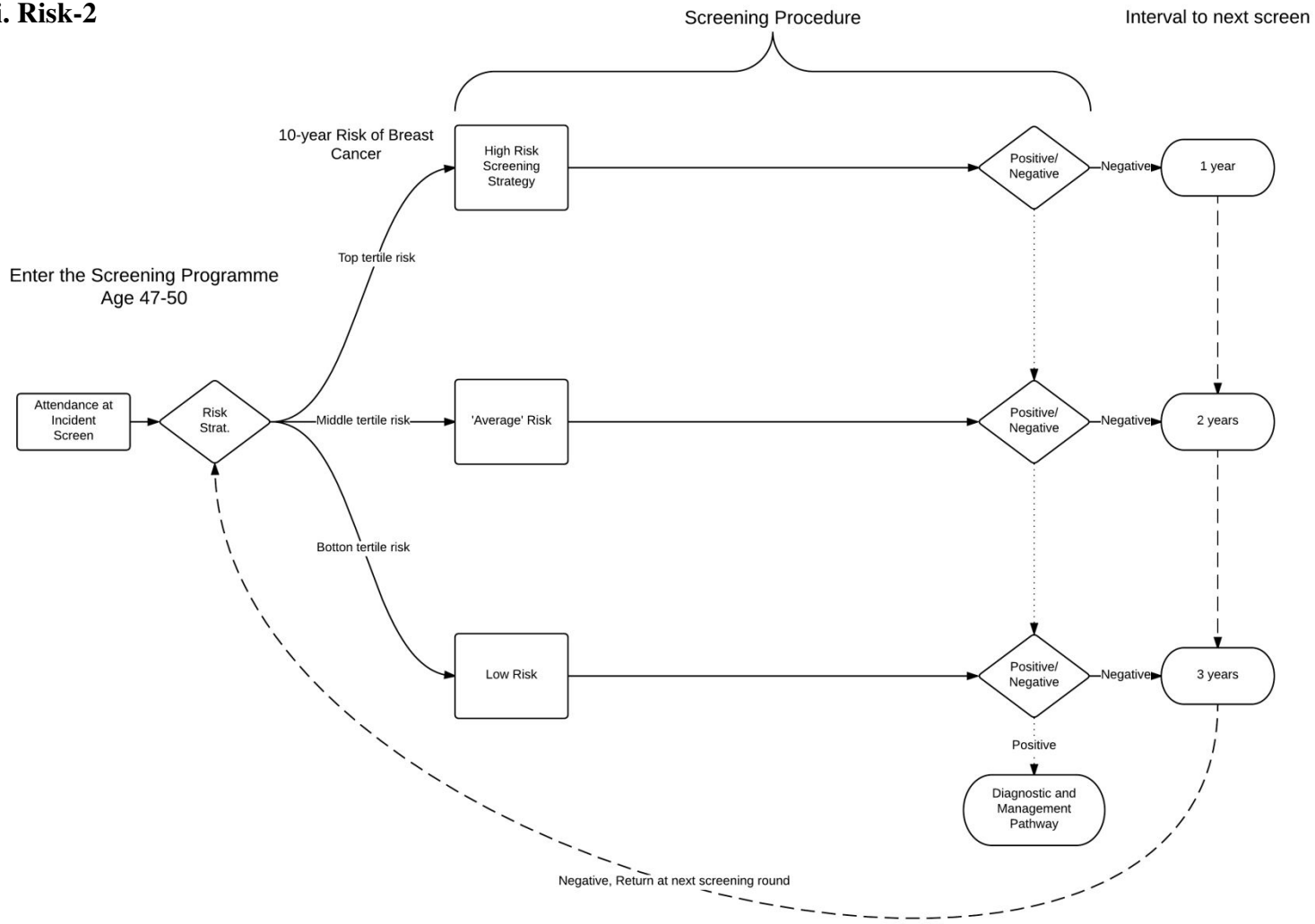
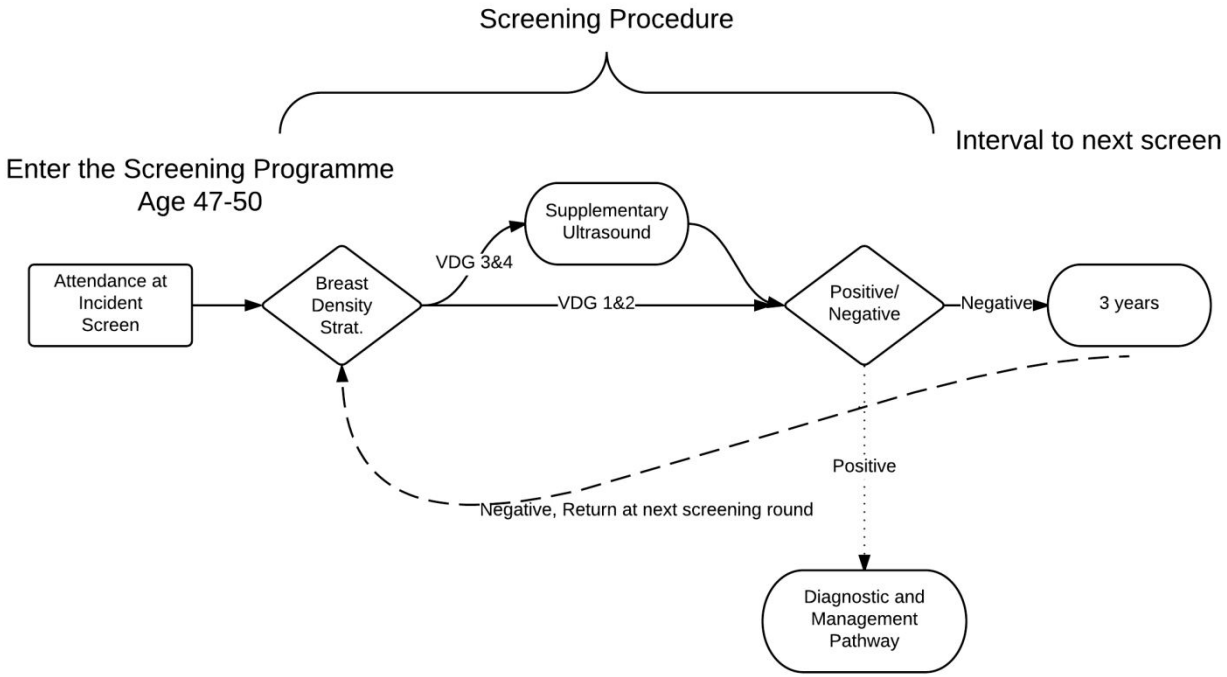
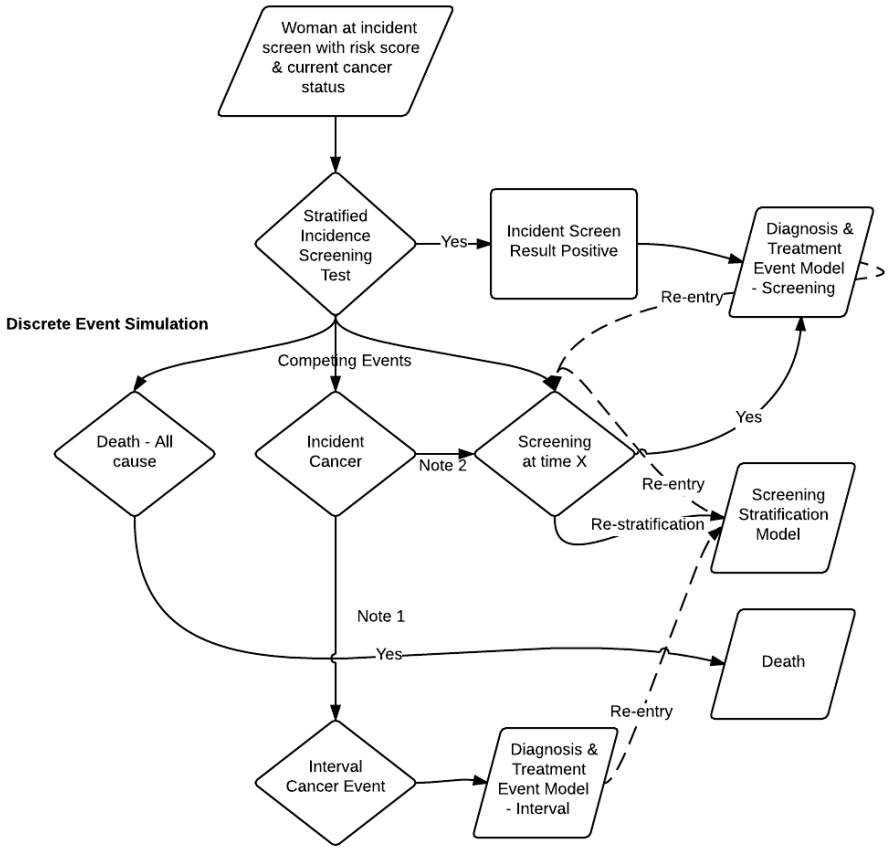


Figure A1.1 – Stratified Programmes
iv. Masking



VDG: Volpara density group

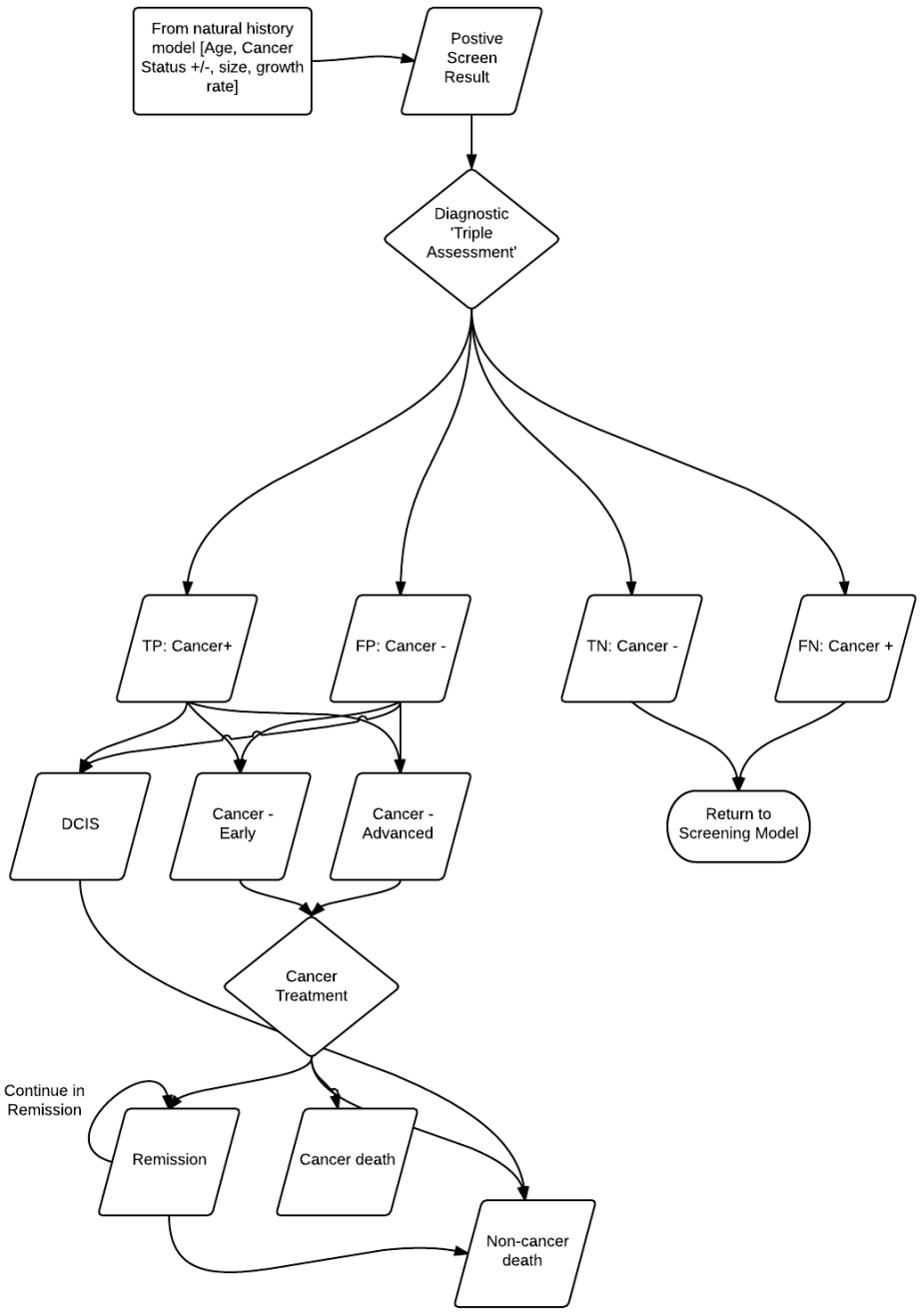
Figure A1.2: Breast cancer natural history and screening



Note 1: A cancer 'natural history' model informs a survival distribution such as Weibull which is used to simulate a time-to-clinical detection for each incident cancer.

Note 2: A individual growth rate from an appropriate distribution is simulated for each individual. A common tumour growth model is then used to calculate tumour size at each screening point. This allows the calculation of positive & negative screening result probabilities based on the relationship between tumour size and screening sensitivity.

Figure A1.3: Diagnosis and treatment



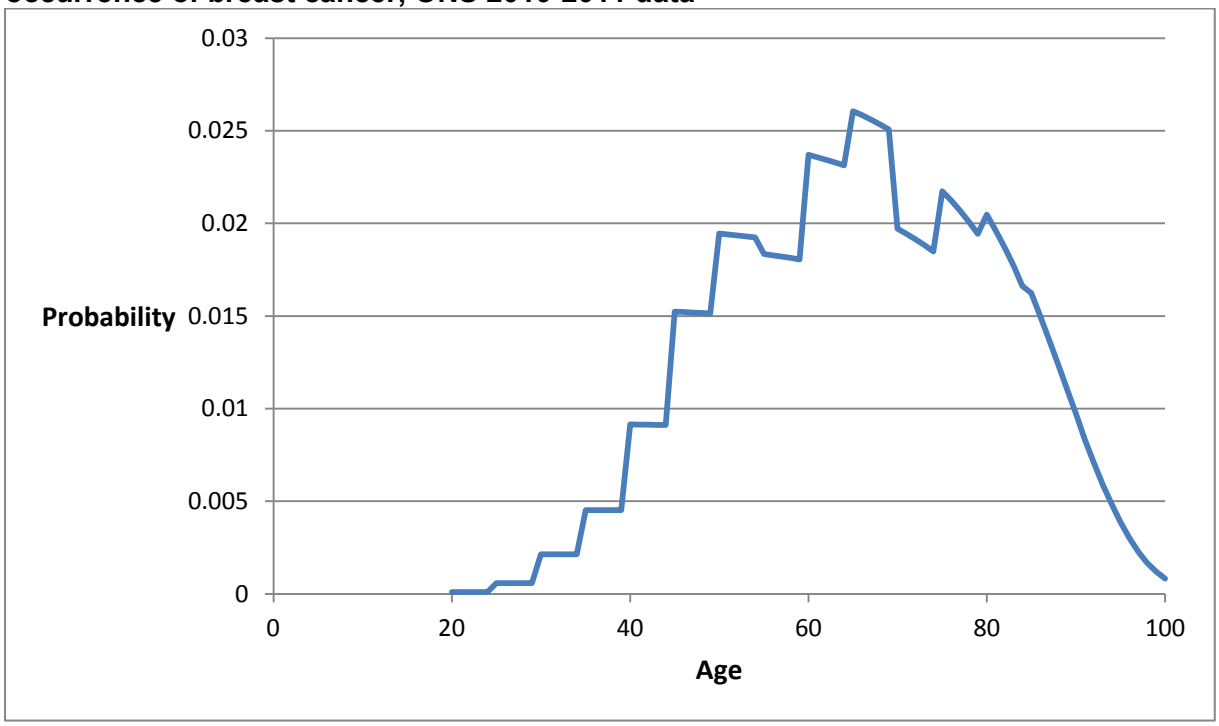
TP: true positive
FP: false positive
DCIS: ductal carcinoma in situ

Appendix 2

Breast cancer incidence

The probability of breast cancer incidence within a five-year age band was calculated as the proportion of all breast cancers that occurred within that age band in the data (see Figure A2.1). Incidence times were allocated uniformly within the five-year age band. To account for the difference between age at diagnosis reported in the data and age at tumour genesis used in the simulation each time was reduced by the mean sojourn time of 2.9 years [5].

Figure A2.1: Probability of breast cancer incidence by age conditional on lifetime occurrence of breast cancer, ONS 2010-2011 data

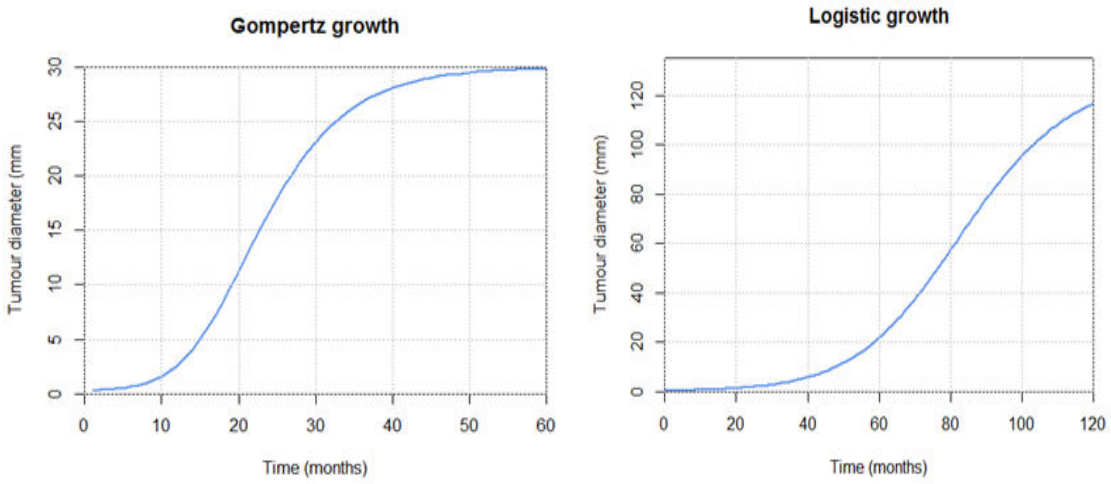


Appendix 3

Tumour growth model functional form choice

Based on the available evidence tumour growth appeared to follow an exponential type curve. However, a simple exponential growth model was deemed to be insufficient to explain observed characteristics of tumour growth; growth slows down as a tumour grows towards maximum possible size and growth rates vary by individual tumour. Using Gompertz and logistic functions allows for an exponential growth that slows towards a maximum size. These functions were therefore deemed to be more appropriate than the simple logistic function for simulating tumour growth. The screening model output will not be sensitive to the choice between Gompertz and logistic because both show near equivalent patterns of growth (see Figure A3.1). The main region of difference close to the maximum tumour size is nearly irrelevant to screening models because almost all tumours are detected by screening or clinical signs before this size.

Figure A3.1: Gompertz and logistic growth functions



An extension of continuous growth natural history models was considered to allow varying tumour growth rates by parameterising the distribution of growth rates in the population. This may be achieved by fitting more complex statistical models to cancer registry data from before and after screening programme initiation. A distribution such as normal or log-normal was assumed for individual tumour growth rates. Random draws from the specified distribution can then be used in the growth simulation. This can more accurately reflect tumour growth across a population of cancer cases and therefore more accurately predict the effect of changing the screening programme. Therefore, this study used Norwegian cancer registry data from before and after the start of a population-based screening programme [5]. A logistic growth function was selected with individual variation in growth rates with a lognormal distribution using parameter values in Table A3.1.

Table A3.1: Cancer growth rate parameters

V_{max} (maximum tumour volume)	$\frac{4}{3}\pi(128/2)^3$ mm^3
V_{cell}(initial tumour size, in theory one cell but in practice may be set to anything)	$\frac{4}{3}\pi(\omega_0/2)^3$mm^3
α_1(mean growth rate)	1.07
α_2 (growth rate variance)	1.31

Invasive cancers

Table A3.2 reports the probability of a tumour of a given size (maximum diameter in mm) being assigned to specific NPI category conditional on the tumour already being an invasive cancer.

Table A1.2 Probability of NPI category membership conditional on tumour size

Size (mm)	NPI I	NPI II	NPI III
1-5	0.76	0.22	0.02
6-10	0.7	0.27	0.02
11-15	0.55	0.43	0.02
16-20	0.4	0.55	0.05
20-30	0.07	0.64	0.29
> 30	0.06	0.5	0.44

Advanced breast cancer

Table A3.3 displays the probability of tumour being at an advanced stage at diagnosis conditional on the size of the tumour (maximum diameter in mm).

Table A3.3 Tumour size and probability of presenting at advanced (TNM: Stage IV) stage

Tumour Size (mm, max diameter)	Probability of presenting at advanced stage
<25	0.046
35	0.087
45	0.110
55	0.127
65	0.143
75+	0.160

Appendix 4

Meta-analysis to inform the parameter for survival with breast cancer

A manuscript is in preparation but the key components of this meta-analysis to inform the parameter for survival with breast cancer is summarised here.

Aim: to identify and assimilate the results of studies that had published estimates of survival stratified by NPI category.

Method: A systematic review and meta-analysis was used to assimilate the data to inform the parameter for survival with breast cancer. The systematic review was limited to studies that included samples of the population of all breast cancer cases (i.e. not limited to screen-detected cancers only). Restricted samples (e.g. only cases receiving a particular therapy) would provide biased estimates of survival if applied to modelling of breast cancer screening which is a population-based intervention.

Two databases (MEDLINE and EMBASE) were searched in May 2015 using an electronic search strategy developed from previous reviews. Identified abstracts and full texts were screened using defined inclusion and exclusion criteria. To be included in the review, the study must have been a prospective cohort study and published as a journal article in English. The study must have directly modelled the relationship between survival or mortality of female breast cancer patients and the NPI-group to which they belong. The study must have provided data on survival and/or mortality either as tables of survivors/events in each year or survival curves presented graphically. Studies that did not meet all of the inclusion criteria were rejected. Exclusion criteria included studies conducted only with a subset of the full population, using patients with recurrent cancer or *ductal carcinoma in situ*.

Results: A total of 649 studies were identified by the electronic search, of which 26 studies were included in the final analysis. Some studies used multiple data series therefore 28 sets of estimates were available from the 26 studies. Half of the identified data series (n=14) used UK data. Nine used data from other European countries. The earliest data used in the included studies came from 1970 [9]. The latest data were from 2010 [10].

Fixed and random effects meta-analysis using the inverse variance method was used to pool study estimates and investigate heterogeneity between the studies. A high degree of heterogeneity between the study estimates was identified, indicating variation in the true survival curves for breast cancer patients in the different study populations. This suggested that using a pooled estimate of survival would be inappropriate.

Using data from the most recent UK study [13] four alternative parametric survival models were estimated. The best fitting model was selected based on Akaike Information Criterion (AIC) and visual inspection of Cox-Snell residuals. See Table A4.1 and Figures A4.1 to A4.3.

All models performed similarly with regard to AIC, with the log-normal and log-logistic having the lowest values in particular groups. The exponential functional form appears to have the best fit considering the plots of Cox-Snell residuals. The exponential model was selected based these considerations. The exponential model has the advantage of having only a single parameter which simplifies the estimation of the PSA. Furthermore, because these survival parameters only apply to a 10-year period in the DES model the additional flexibility of the two parameter models is a less important consideration.

Table A4.1 – Parametric survival models – AIC statistics

Group	Exponential	Weibull	Log-normal	Log-logistic
NPI 1	902.703	890.9548	886.5255	890.058
NPI 2	793.8253	784.8852	775.8036	782.7285
NPI3	261.6141	258.3425	256.857	257.9818
All NPI groups	1983.964	1957.3	1942.372	1954.127

Figure A4.1 – Cox-Snell residuals plot – NPI group 1

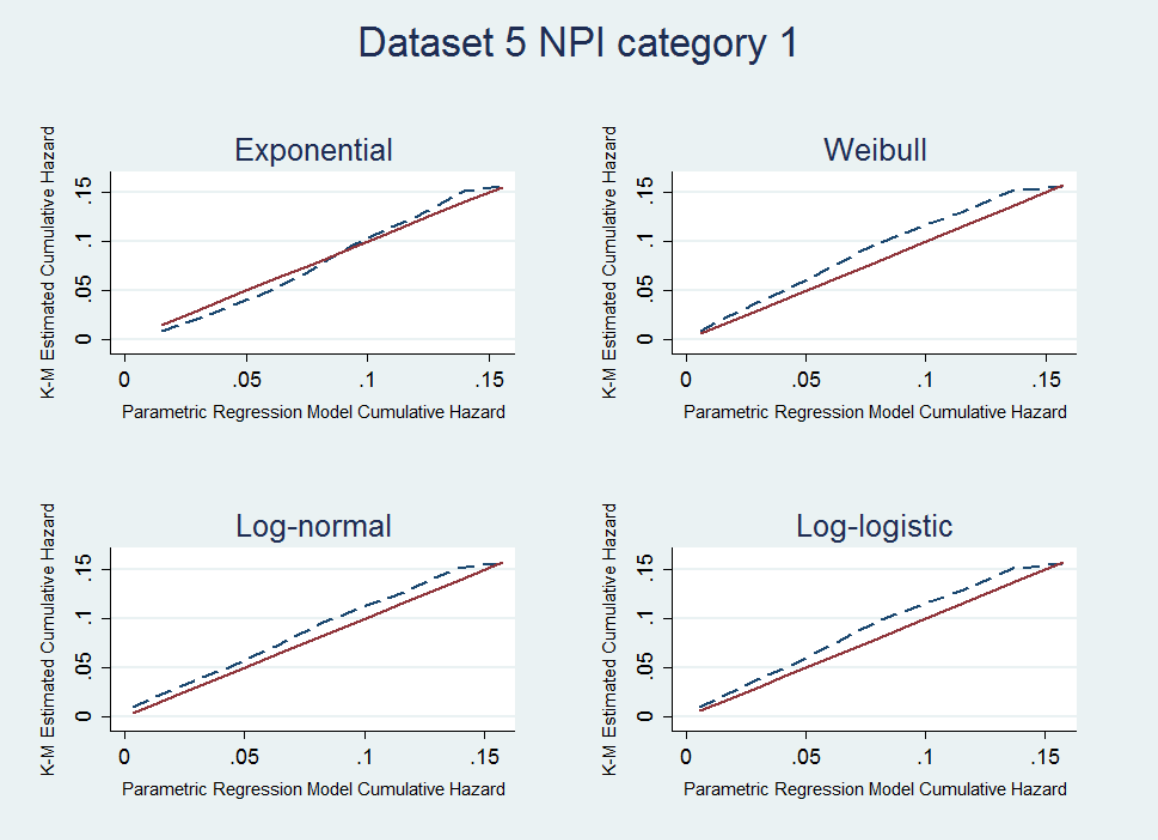


Figure A4.2 – Cox-Snell residuals plot – NPI group 2

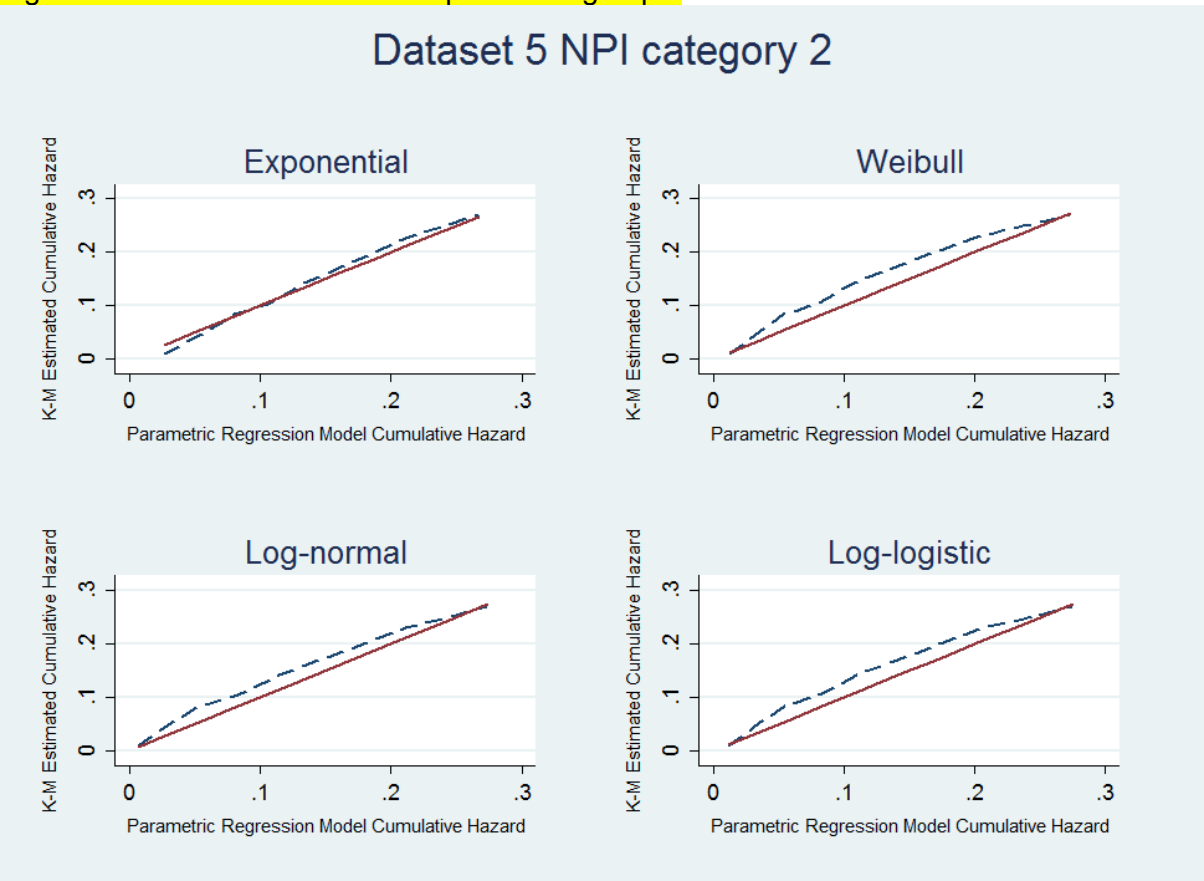
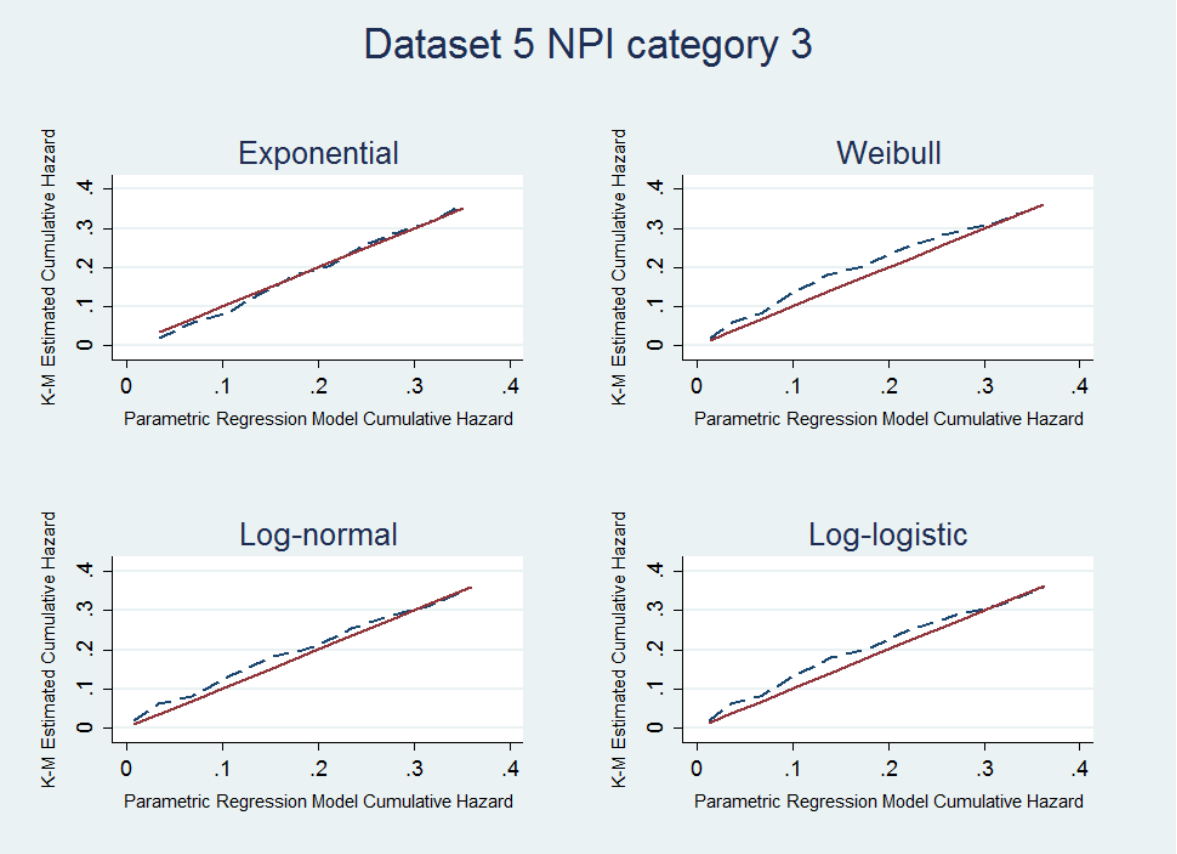


Figure A4.3 Cox-Snell residuals plot – NPI group 3



Appendix 5

Visual assessment of Monte Carlo simulation error

The histograms (Figure A5.1 to A5.6) show a visual assessment of the inherent error in the Monte Carlo simulations. The histograms report the incremental costs and incremental QALYs, for the Risk 1 stratified-NBSP compared with the current-NBSP, using sample sizes: 1 million; 10 million and 100 million women per alternative. These results show that at the point at which100 million is reached the sample size errors become relatively small, which gives confidence that the model has sufficiently converged.

Figure A5.1: Incremental QALYs for a sample of 1 million women per alternative

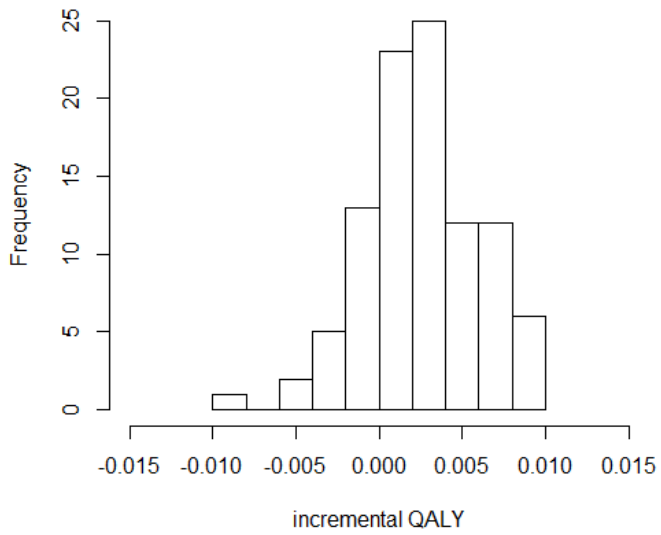


Figure A5.2: Incremental QALYs for a sample of 10 million women per alternative

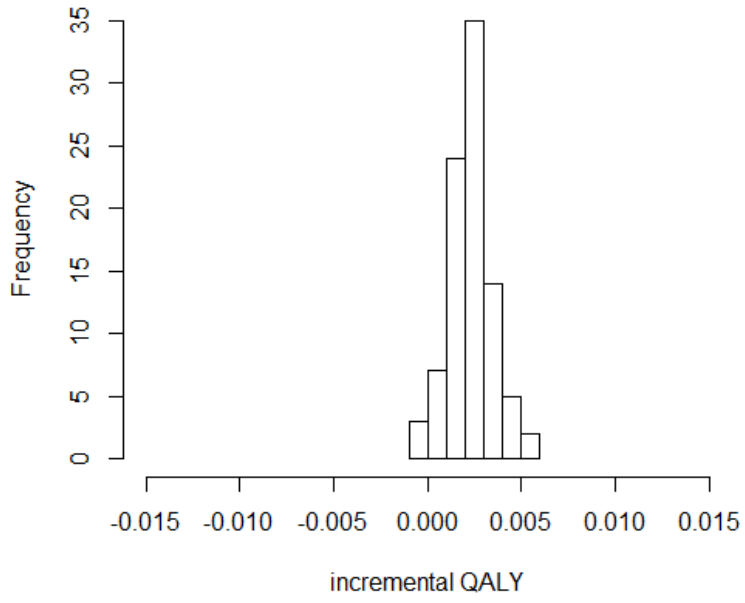


Figure A5.3: Incremental QALYs for a sample of 100 million women per alternative

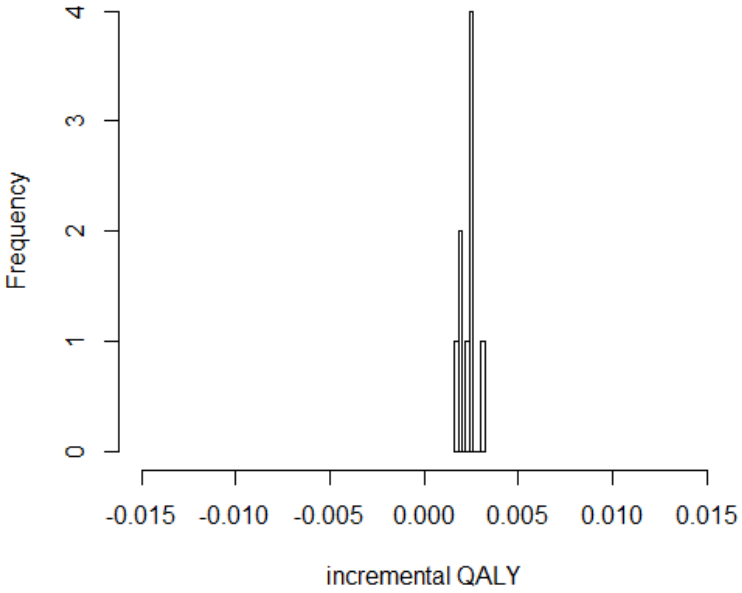


Figure A5.4: Incremental costs for a sample of 1 million women per alternative

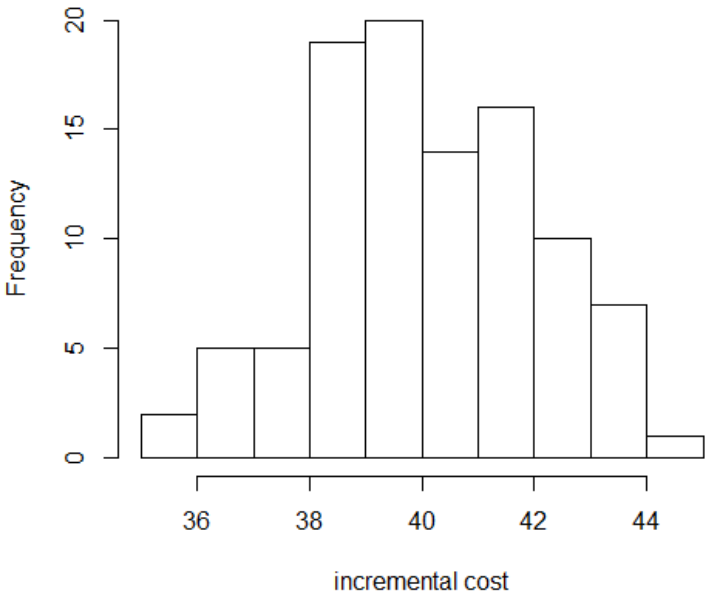


Figure A5.5: Incremental costs for a sample of 10 million women per alternative

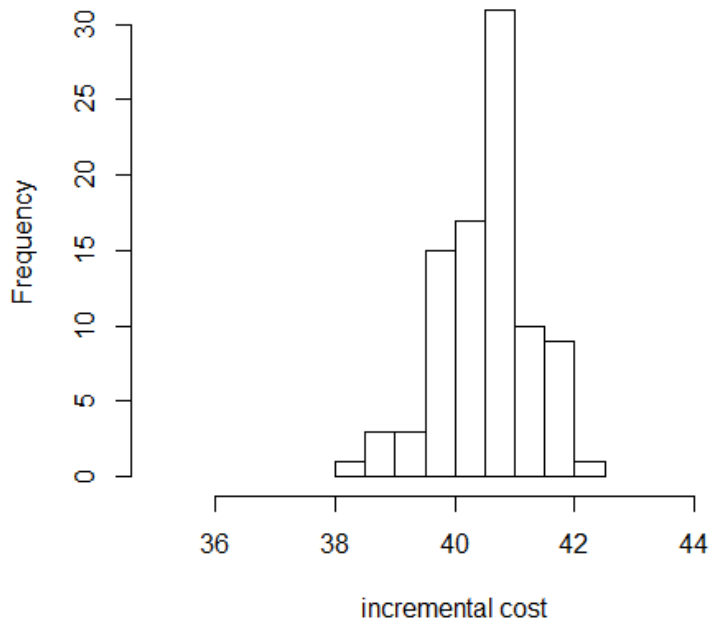
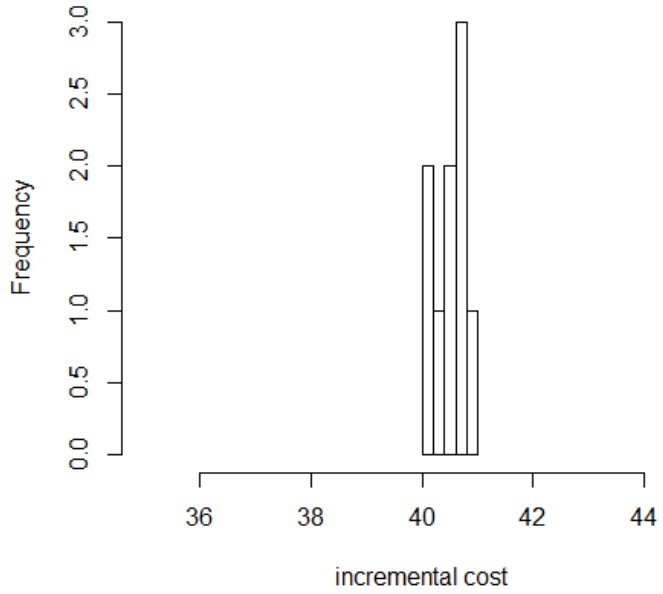


Figure A5.6: Incremental costs for a sample of 100 million women per alternative



Appendix 6

One-way sensitivity analysis

A series of one-way sensitivity analyses (OWSA) were used to identify the key drivers of the relative cost-effectiveness of the example stratified-NBSP. The selection of the parameters and model input values to include in the sensitivity analysis was informed by a logical assessment of the key assumptions made when populating the model. All parameters in the OWSA were varied by four discrete values representing a decrease (-20% and -10%) and increase (+10% and +20%) of their base case values. The parameters selected for inclusion in each of the one-way sensitivity analyses are shown in Table A6.1.

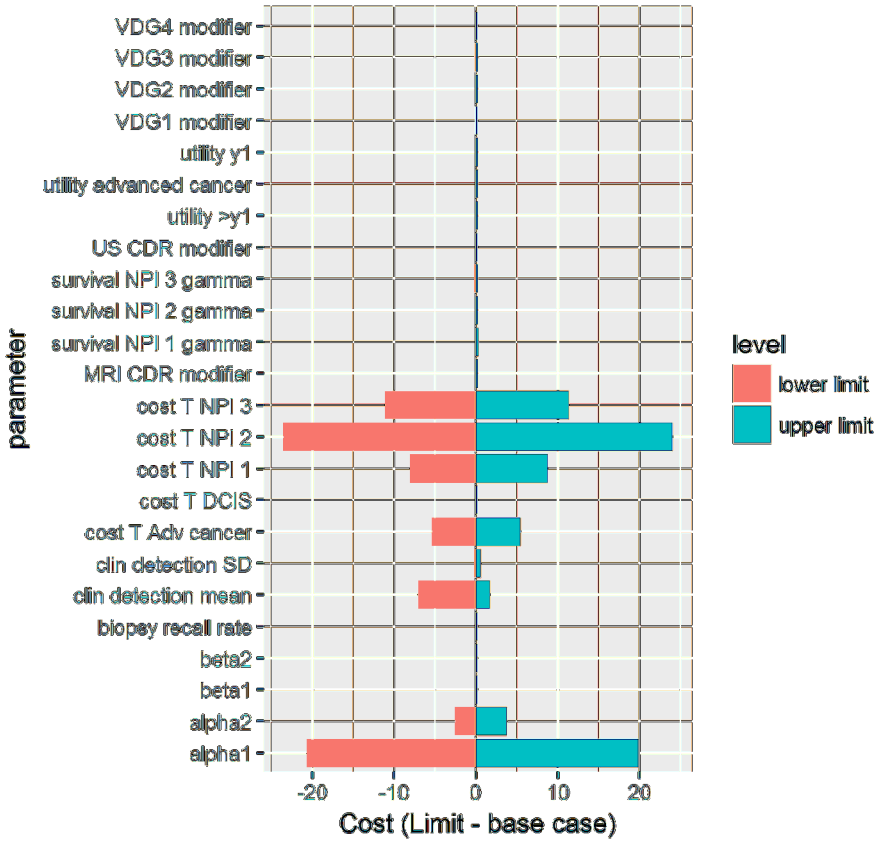
In addition, three further scenario analyses were conducted to explore the impact of changing the assumptions in the value of: biopsy recall rate with ultrasound (assumed 2.4% in basecase); the unit cost of the stratification process (£10.57 per woman in basecase); inflation index used for the published treatment costs.

Table A6.1: Parameters used in the one-way sensitivity analysis

Parameter	Base case value
Natural history parameters:	
α_1 growth rate	1.07
α_2 growth rate	1.31
β_1 mammographic sensitivity	1.47
β_2 mammographic sensitivity	6.51
VDG sensitivity modifiers:	
Sensitivity VDG1	85.0%
Sensitivity VDG2	77.6%
Sensitivity VDG3	69.0%
Sensitivity VDG4	58.6%
Supplemental screening parameters:	
VDG3/4 incremental cancers detected with supplemental US	3 per 1000 exams
VDG3/4 incremental cancers detected with supplemental MRI	5 per 1000 exams
Survival parameters:	
γ NPI 1	-5.413
γ NPI 2	-4.023
γ NPI 3	-2.465
Utility parameters:	
Early breast cancer – first year	0.696
Early breast cancer – subsequent years	0.779
Advanced breast cancer	0.685
Treatment costs:	
NPI 1 treatment (mean)	£11630
NPI 2 treatment (mean)	£12978
NPI 3 treatment (mean)	£15405
Advanced cancer (mean)	£23449
Other parameters:	
Recall biopsy rate	2.4
Cancer size at clinical detection (mean)	6.5 doublings (2.62mm)
Cancer size at clinical detection (standard deviation)	0.535 doublings

The following Figures (Figure A5.1 to A5.12) show the results of the impact of each one-way sensitivity analysis using tornado plots with separate plots reported for each alternative screening programme and costs and QALYs separately. Values on the x-axis are the difference between the upper or lower limit value of costs or QALYs and the respective value in the base case.

Figure A6.1: No screening - costs

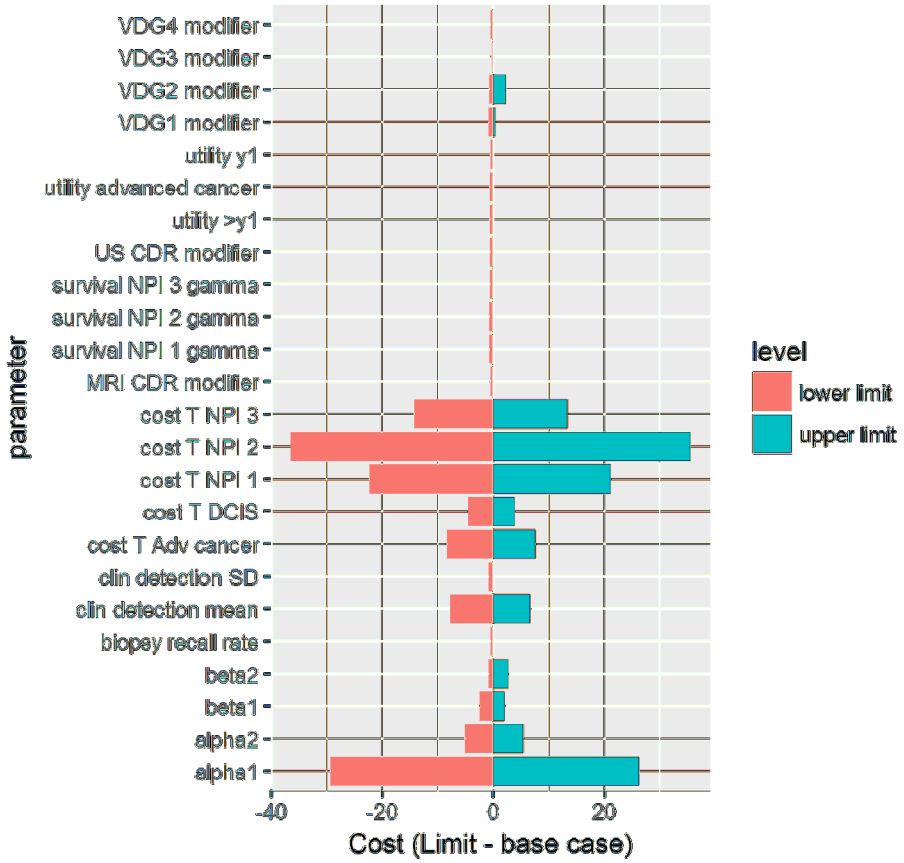


Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.2: No screening - QALYs

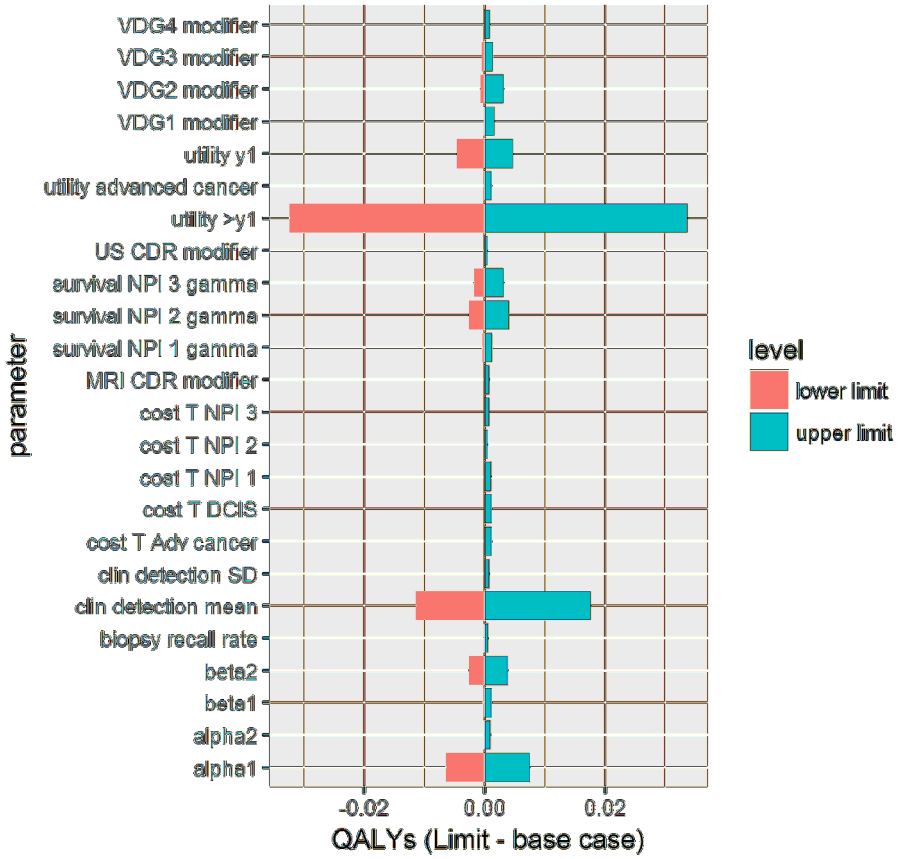
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.3: current NBSP - Costs



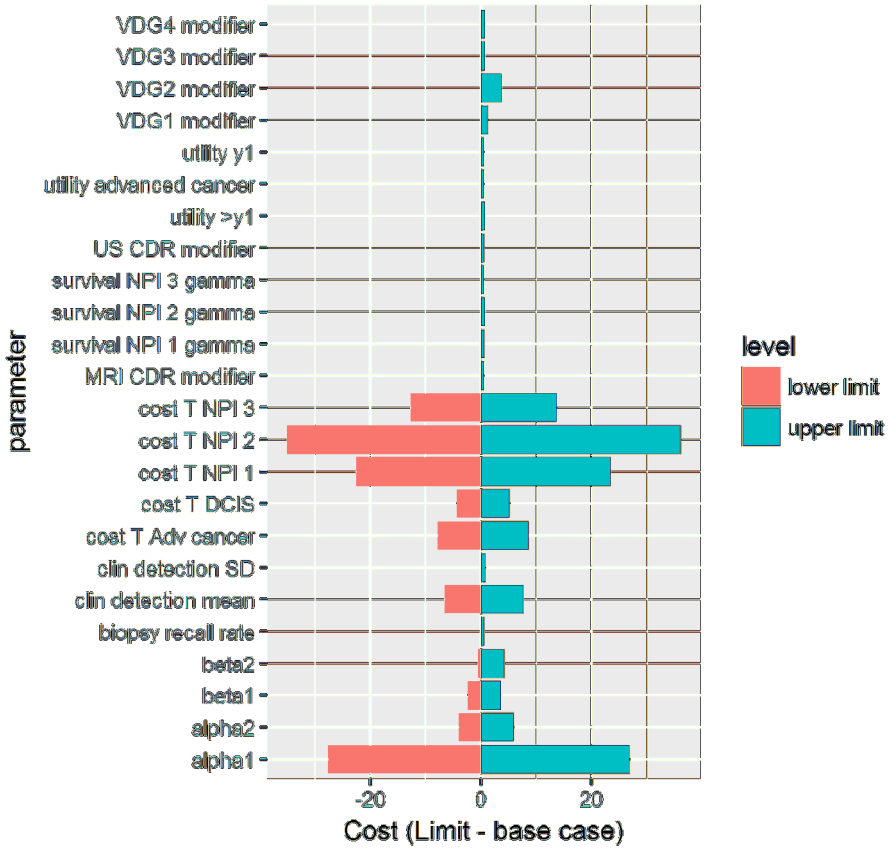
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.4: current NBSP - QALYs



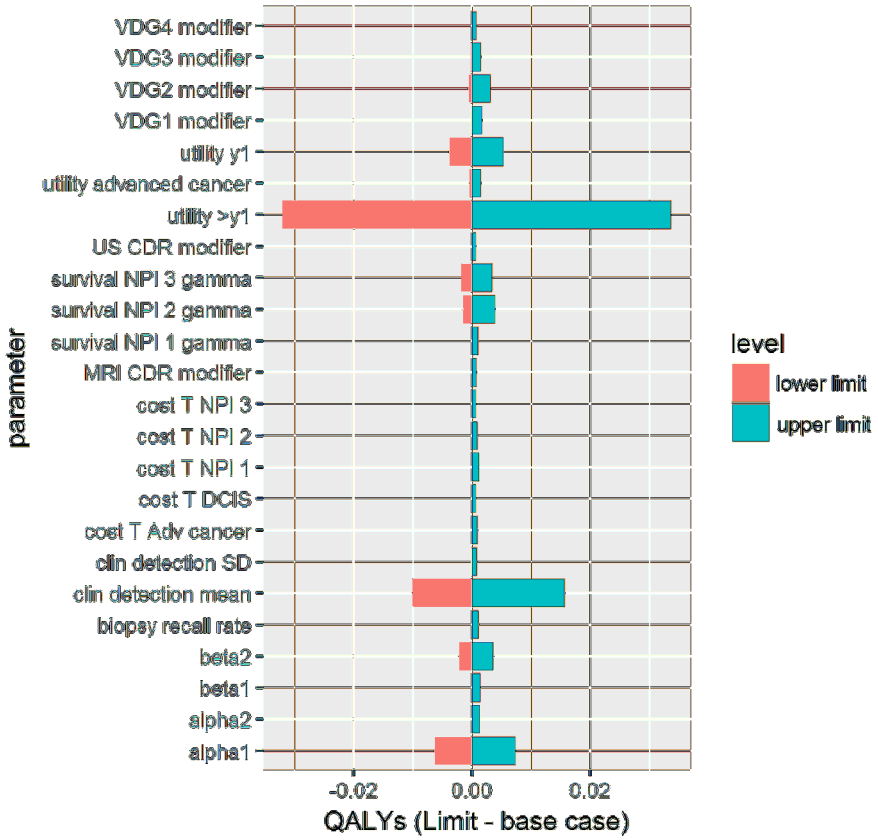
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.5: Risk 1 stratified NBSP - Costs



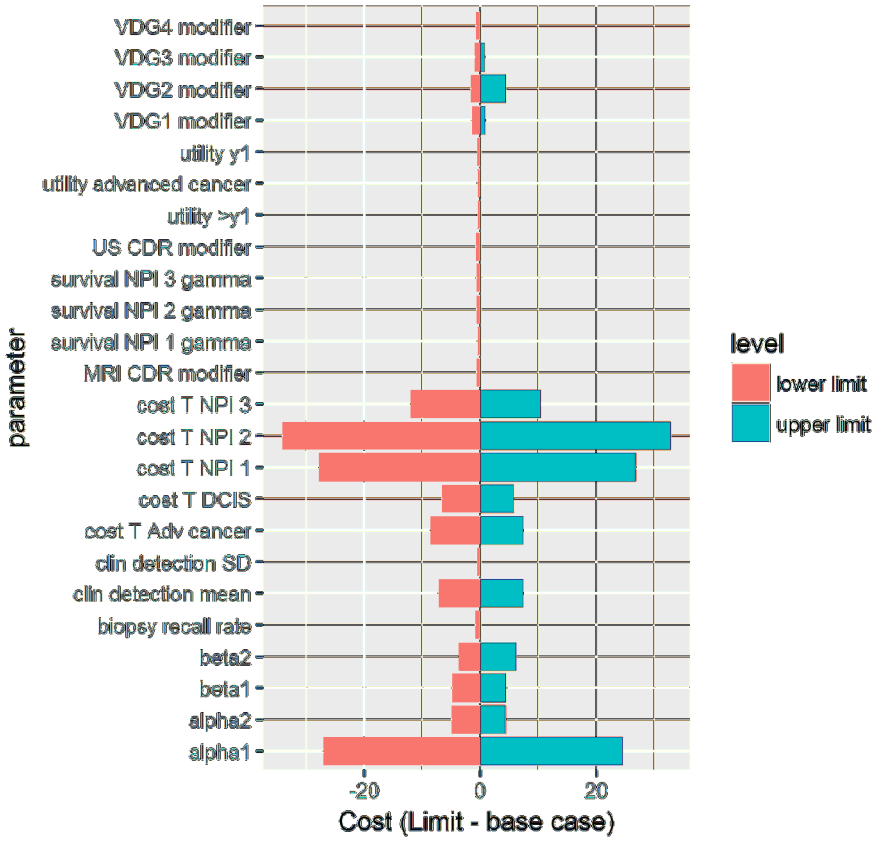
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.6: Risk 1stratified NBSP - QALYs



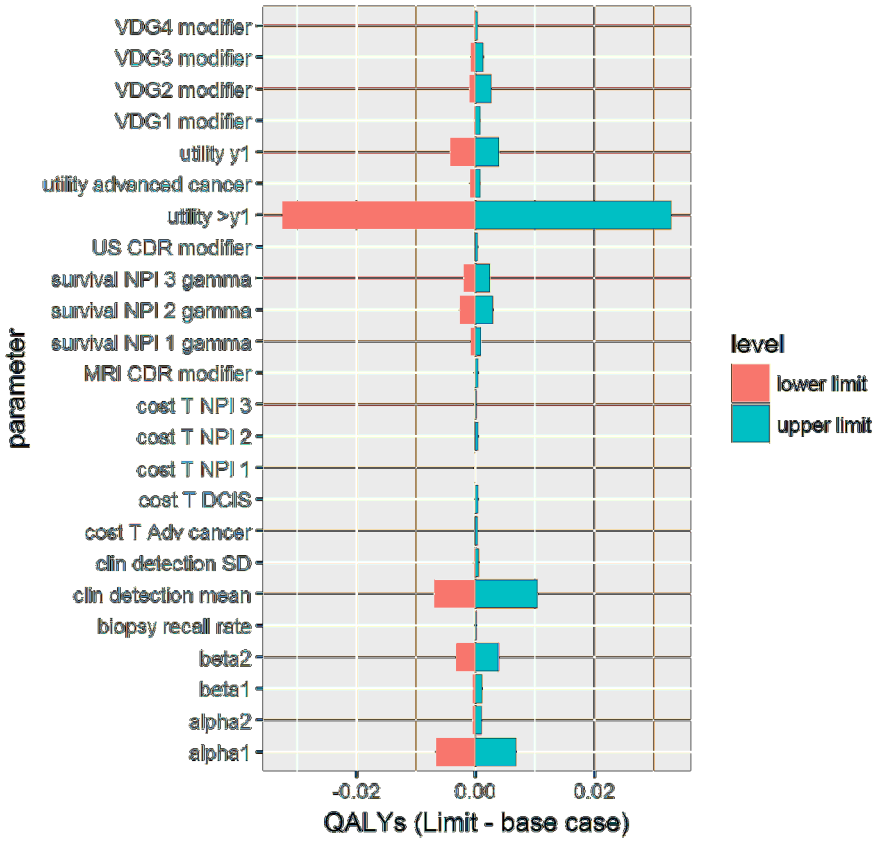
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.7: Risk 2 stratified NBSP - Costs



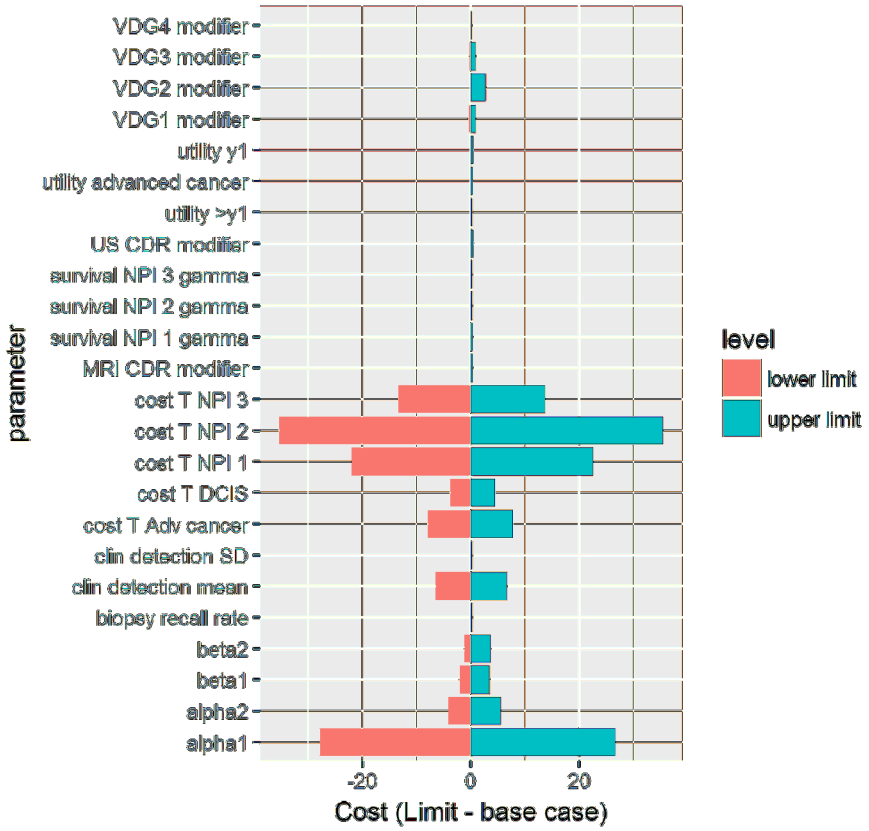
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.8: Risk 2 stratified NBSP - QALYs



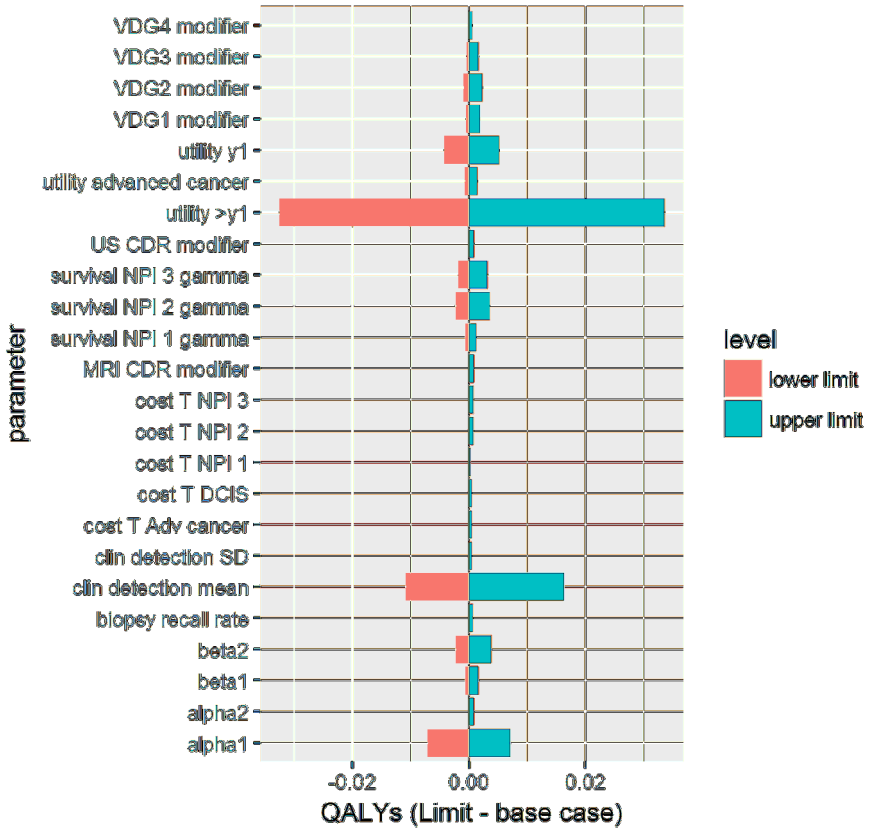
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.9: Masking stratified NBSP - Costs



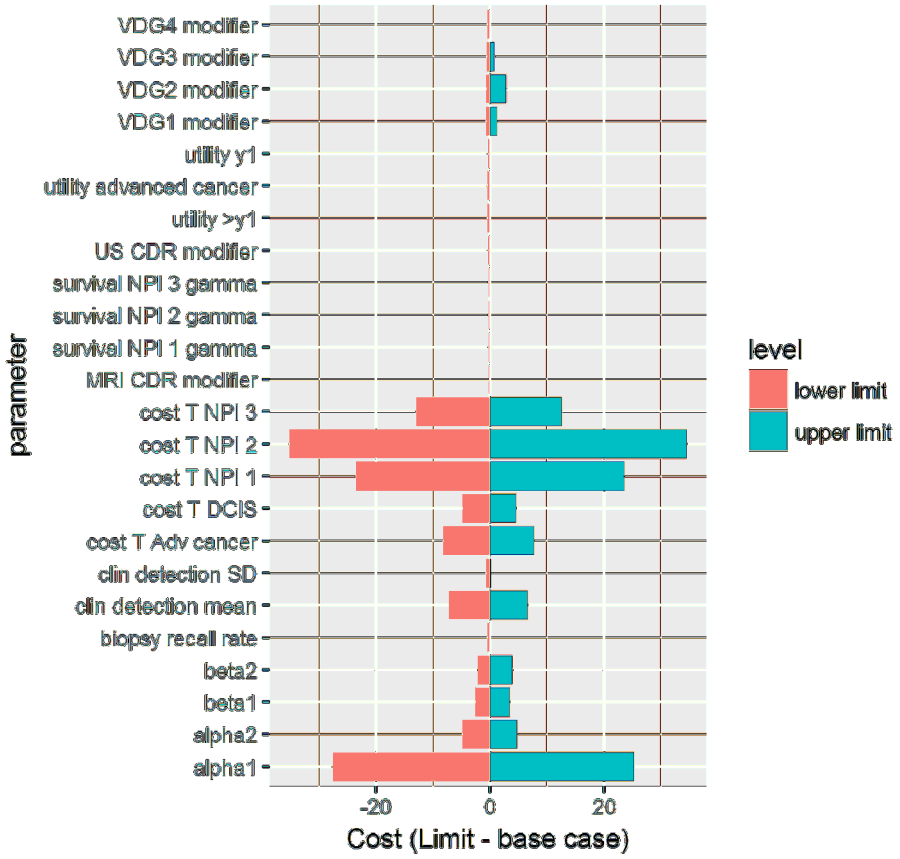
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.10: Masking stratified NBSP - QALYs



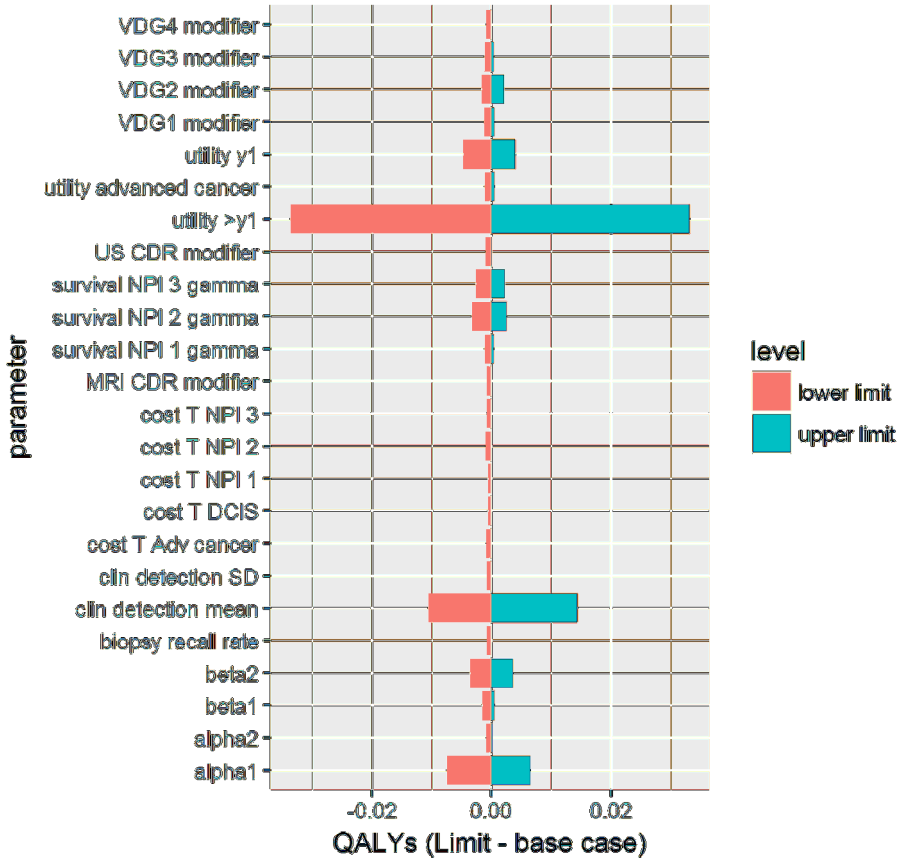
Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.11: Risk 1 & Masking stratified NBSP - Costs



Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Figure A6.12: Risk 1 & Masking stratified NBSP - QALYs



Legend definition: four discrete values representing lower limit (-20% and -10%) & upper limit (+10% and +20%)

Table A6.2. Incremental cost effectiveness ratios (ICERs) for the results of the one-way sensitivity analysis

Parameter	Screening strategy	Incremental cost per QALY gained (£ per QALY)*					
		Basecase parameter values		Change in basecase parameter value			
				Decreased by 20%		Increased by 20%	
		Screening strategy vs no screening**	Screening strategy vs current-NBSP**	Screening strategy vs no screening	Screening strategy vs current-NBSP	Screening strategy vs no screening	Screening strategy vs current-NBSP
Mammographic Sensitivity Beta 1	No screening	NA	NA	NA	NA	NA	NA
	Current-NBSP	23197	NA	24066	NA	22650	NA
	Risk 1	22413	16689	23282	17465	22265	19045
	Risk 2	23435	23924	23646	22846	22825	23180
	Masking	**30772**	**212947**	**32080**	**243868**	29014	**109919**
	Risk 1 & Masking	**30532**	**75254**	**31805**	**80015**	29401	**66881**
Mammographic Sensitivity Beta 2	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	19229	NA	27743	NA
	Risk 1	22413	16689	19576	23880	26310	17223
	Risk 2	23435	23924	20402	23191	26471	24229
	Masking	**30772**	**212947**	25404	**163981**	**35426**	**130514**
	Risk 1 & Masking	**30532**	**75254**	25156	**60042**	**35391**	**73943**
Growth rate Alpha 1	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22194	NA	24063	NA
	Risk 1	22413	16689	22245	22754	23503	18948
	Risk 2	23435	23924	23007	24755	23666	22894
	Masking	**30772**	**212947**	29924	**268952**	**32543**	**540516**
	Risk 1 & Masking	**30532**	**75254**	29454	**73345**	**31171**	**72139**
Growth rate Alpha 2	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	21826	NA	23753	NA
	Risk 1	22413	16689	20986	15057	23331	19822
	Risk 2	23435	23924	21710	21481	24154	24988
	Masking	**30772**	**212947**	28310	**131405**	**30179**	**101503**
	Risk 1 & Masking	**30532**	**75254**	29068	**78482**	**31269**	**76551**
VDG modifier 1	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23549	NA	21425	NA
	Risk 1	22413	16689	23264	20736	21266	19790
	Risk 2	23435	23924	23437	23218	22451	24822
	Masking	**30772**	**212947**	**31390**	**242047**	27789	**124975**
	Risk 1 & Masking	**30532**	**75254**	**30980**	**76068**	28625	**77558**
VDG modifier 2	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	24557	NA	19649	NA
	Risk 1	22413	16689	24029	19753	19770	21076

	Risk 2	23435	23924	24197	23508	21025	24417
	Masking	**30772**	**212947**	**32584**	**222788**	27030	**2743678**
	Risk 1 & Masking	**30532**	**75254**	**31779**	**70907**	26407	**75345**
VDG modifier 3	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23819	NA	21594	NA
	Risk 1	22413	16689	23215	18464	21401	19634
	Risk 2	23435	23924	23622	23238	22073	23090
	Masking	**30772**	**212947**	**31041**	**153405**	28033	**126779**
	Risk 1 & Masking	**30532**	**75254**	**30717**	**67680**	28684	**74620**
VDG modifier 4	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23060	NA	21951	NA
	Risk 1	22413	16689	22745	19991	22104	23770
	Risk 2	23435	23924	23549	24590	22657	24207
	Masking	**30772**	**212947**	**30510**	**200944**	29335	**249898**
	Risk 1 & Masking	**30532**	**75254**	**30054**	**70189**	28923	**71714**
US cancer detection rate (CDR) modifier	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23570	NA	22936	NA
	Risk 1	22413	16689	22957	18191	22749	21027
	Risk 2	23435	23924	23377	23000	23258	23925
	Masking	**30772**	**212947**	**30183**	**114152**	**29541**	**119283**
	Risk 1 & Masking	**30532**	**75254**	**31043**	**76907**	**29370**	**62064**
MRI cancer detection rate (CDR) modifier	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23359	NA	22025	NA
	Risk 1	22413	16689	22276	15189	22487	28554
	Risk 2	23435	23924	23161	22777	22452	23354
	Masking	**30772**	**212947**	**30134**	**124262**	29364	**231631**
	Risk 1 & Masking	**30532**	**75254**	29687	**60418**	29618	**84178**
Survival NPI 1 gamma	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22366	NA	23790	NA
	Risk 1	22413	16689	22475	23647	22553	14850
	Risk 2	23435	23924	22748	23549	23832	23916
	Masking	**30772**	**212947**	29299	**154537**	**31799**	**263202**
	Risk 1 & Masking	**30532**	**75254**	29324	**70617**	**30794**	**68872**
Survival NPI 2 gamma	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22593	NA	21840	NA
	Risk 1	22413	16689	22598	22651	20637	13231
	Risk 2	23435	23924	23591	25862	22133	22740
	Masking	**30772**	**212947**	**30445**	**328442**	28201	**118834**
	Risk 1 & Masking	**30532**	**75254**	**30547**	**89852**	28448	**65952**
Survival NPI 3 gamma	No screening	NA	NA	NA	NA	NA	NA

	NBSP	23197	NA	25157	NA	21400	NA
	Risk 1	22413	16689	24304	18143	21419	21610
	Risk 2	23435	23924	25033	24789	22043	23445
	Masking	**30772**	**212947**	**32803**	**159497**	28257	**174156**
	Risk 1 & Masking	**30532**	**75254**	**32486**	**71560**	28246	**70730**
Utility year 1	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23997	NA	23351	NA
	Risk 1	22413	16689	22724	14799	22385	15801
	Risk 2	23435	23924	23250	21898	23777	24674
	Masking	**30772**	**212947**	**30662**	**111372**	29778	**106547**
	Risk 1 & Masking	**30532**	**75254**	**30087**	**57446**	**30268**	**68346**
Utility after year 1	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22185	NA	23173	NA
	Risk 1	22413	16689	21619	17192	22981	21214
	Risk 2	23435	23924	22378	22771	23736	24941
	Masking	**30772**	**212947**	29270	**177570**	**30564**	**183390**
	Risk 1 & Masking	**30532**	**75254**	29607	**79616**	29759	**63746**
Utility advanced cancer	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22917	NA	22898	NA
	Risk 1	22413	16689	22877	22481	22226	17139
	Risk 2	23435	23924	23436	24542	23089	23477
	Masking	30772	212947	30874	333266	29560	124029
	Risk 1 & Masking	30532	75254	30067	72733	29558	65154
Cost of treatment for DCIS	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23344	NA	23303	NA
	Risk 1	22413	16689	22649	17405	23260	22833
	Risk 2	23435	23924	22993	22325	23843	24998
	Masking	**30772**	**212947**	**30688**	**165309**	**30668**	**186940**
	Risk 1 & Masking	**30532**	**75254**	**30119**	**65660**	**30361**	**71286**
Cost of treatment for NPI 1	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	21855	NA	23159	NA
	Risk 1	22413	16689	21128	15850	23190	23510
	Risk 2	23435	23924	22515	23945	23841	25332
	Masking	**30772**	**212947**	29524	**267708**	**31082**	**387016**
	Risk 1 & Masking	**30532**	**75254**	29006	**72258**	**30650**	**80915**
Cost of treatment for NPI 2	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22293	NA	23896	NA
	Risk 1	22413	16689	22047	19927	23105	17176
	Risk 2	23435	23924	22976	24405	23044	21455
	Masking	**30772**	**212947**	29788	**200046**	**30750**	**137414**
	Risk 1 & Masking	**30532**	**75254**	29059	**64678**	**30916**	**72154**

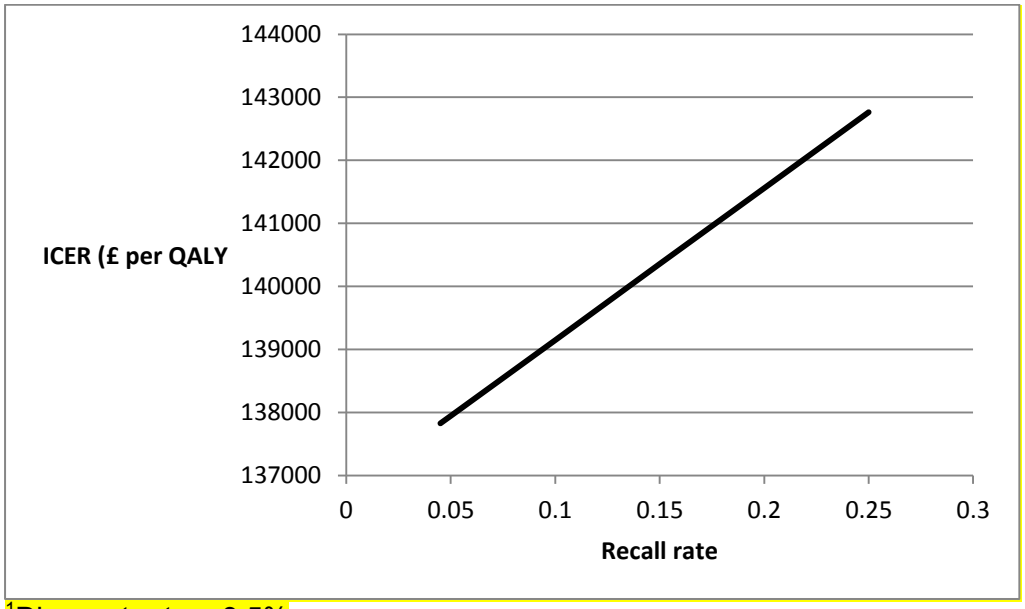
Cost of treatment for NPI 3	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22612	NA	23866	NA
	Risk 1	22413	16689	22805	24905	23237	18286
	Risk 2	23435	23924	23627	25904	23385	22464
	Masking	**30772**	**212947**	**30963**	**766499**	**30680**	**124538**
	Risk 1 & Masking	**30532**	**75254**	**30535**	**87568**	**31121**	**73711**
Cost of treatment for advanced cancer	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22846	NA	22618	NA
	Risk 1	22413	16689	22823	22590	22320	19694
	Risk 2	23435	23924	23293	24232	23064	24014
	Masking	**30772**	**212947**	**30278**	**198308**	**30268**	**285156**
	Risk 1 & Masking	**30532**	**75254**	**30658**	**84864**	**30401**	**87241**
Biopsy recall rate	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	22930	NA	23377	NA
	Risk 1	22413	16689	22013	15684	22144	14447
	Risk 2	23435	23924	23203	23768	23046	22410
	Masking	**30772**	**212947**	**30436**	**213347**	29832	**107946**
	Risk 1 & Masking	**30532**	**75254**	**29941**	**70458**	**30097**	**65547**
Clinical detection mean	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	**121171**	NA	11786	NA
	Risk 1	22413	16689	**131331**	**892973**	11823	12210
	Risk 2	23435	23924	**116953**	**109301**	12756	15248
	Masking	**30772**	**212947**	**182879**	**-512287**[#]	15630	**104161**
	Risk 1 & Masking	**30532**	**75254**	**152188**	**297453**	15883	**45672**
Clinical detection SD	No screening	NA	NA	NA	NA	NA	NA
	NBSP	23197	NA	23174	NA	23086	NA
	Risk 1	22413	16689	23016	21527	22406	17307
	Risk 2	23435	23924	23367	23762	22945	22671
	Masking	**30772**	**212947**	**30287**	**153932**	**30296**	**165263**
	Risk 1 & Masking	**30532**	**75254**	**30235**	**70632**	29784	**65417**

*ICERs exceeding £20,000 per QALY are underlined; ICERs exceeding £30,000 per QALY are bold
**Base case ICERs in this table are based on rounded disaggregated costs and QALYs therefore there are small differences with base case ICERs in main text of the manuscript.
[#]QALY loss

The following sections describe the results from the three additional scenario analyses.

Biopsy recall rate with ultrasound
The recall rate for supplemental screening with ultrasound may be higher than for mammography in the current-NBSP. To explore the sensitivity of the Masking-1 NBSP to this parameter we varied the recall rate for ultrasound screening from 4.5% (mammography rate) to 25%. The effect on the ICER of the masking 1 strategy compared to current screening can be seen in Figure A6.13. The rate of recall for ultrasound has only a moderate effect on the ICER for the range examined.

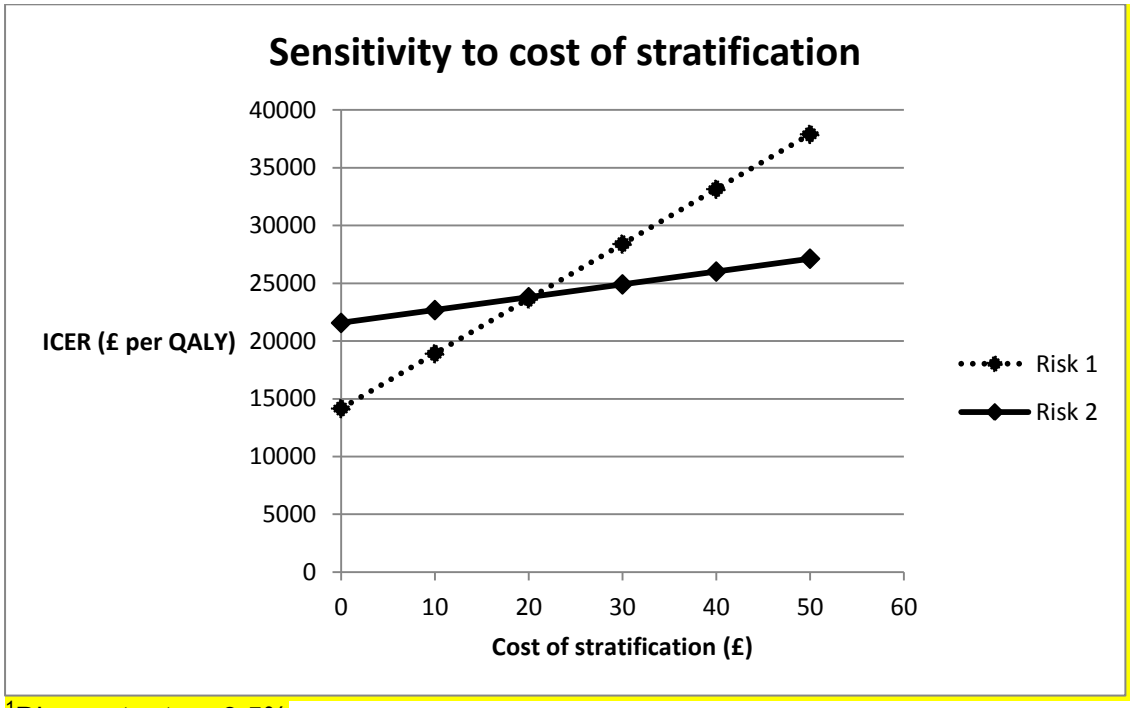
Figure A6.13: One-way sensitivity analysis on biopsy recall rate for US, Masking stratified NBSP vs current NBSP[1]



[1]Discount rate = 3.5%

Cost of stratification
The cost of implementing the stratification approach was not easily quantified in the basecase analysis because of the potential for differences in the approach and a lack of actual resource use data. It was assumed that this cost could vary substantially depending on the details of the procedure used, e.g. a letter with risk information only or a consultation with a clinician. Instead of trying to assume the procedure and associated cost we investigated applying an average cost per woman of stratification in the range of £0 to £50. The ICERs for Risk-1 NBSP and Risk-2 NBSP were compared with the current UK-NBSP using varying stratification costs. The results are displayed in Figure A6.14. This scenario analysis demonstrates that if stratification has a high cost such as £50 per woman the Risk 1 stratified screening may not be considered cost-effective. Risk 2 stratified screening remained a cost-effective alternative to current screening even at a £50 per woman cost of stratification.

Figure A6.14: Results of the one-way sensitivity analysis on assumed cost of stratification using Risk-1 or Risk-2 compared with the current NBSP[1]



[1]Discount rate = 3.5%

Treatment cost inflation factors

Treatment cost estimates were only available from a dated source. The basecase analysis chose to use the most commonly used method (the CPI) to inflate the cost estimate to reflect the 2015 price year. However, the method used to inflate and adjust to present values may be influential. In particular, there was concern that relatively new treatments (adjuvant chemotherapy agents) for breast cancer have greatly changed the treatment costs in this time period. As an alternative to using the CPI as in the base case, treatment costs have been adjusted by using a 10% per year inflation adjustment. It was not possible to identify a source that was specific to breast cancer and also the time period, or a source which could break down costs into prognostic groups, therefore using this approach was the only feasible way of exploring the impact of using a different inflation rate. Results are shown in Table A6.3. This scenario analysis showed that using this inflation figure increased the individual ICERs. This meant that for the current-NBSP, risk-1 stratified NBSP and risk-2 stratified NBSP compared with no screening the ICERs moved close to the £30,000 per QALY threshold and for masking and risk-1 and masking stratified NBSP compared with no screening the ICERs were above the £30,000 per QALY threshold. When compared with the current NBSP, the size of the ICERs was not increased substantially with the exception of the masking NBSP.

Table A6.3: Scenario analysis using inflation adjustment (10%) for treatment costs

Screening Programme	**QALYs (3.5% discount rate)**	**Cost (£; 2015; 3.5% discount rate)**	**Scenario analysis Incremental cost-effectiveness ratio (ICER)* £ per QALY**		**Basecase analysis Incremental cost-effectiveness ratio (ICER £ per QALY**	
			Versus no screening (3.5% discount rate)	**versus current-NBSP (3.5% discount rate)**	**Versus no screening (3.5% discount rate)**	**versu curren NBSP (3.5% discou rate)**
No screening	17.692	430	N/A	N/A	N/A	N/A
Current UK-NBSP	17.710	952	29419	N/A	23197	N/A
Risk-1	17.712	993	28050	17506	22413	1668
Risk-2	17.718	1154	27516	23565	23435	2392
Masking	17.711	1107	**35962**	**144276**	**30772**	**21294**
Risk-1 and Masking	17.712	1169	**35873**	**76144**	**30532**	**75254**

*ICERs exceeding £20,000 per QALY are underlined; ICERs exceeding £30,000 per QALY are bold

Appendix 7

Probabilistic Sensitivity Analysis (PSA)

The PSA was conducted using the Monte Carlo simulation method, taking 1000 random draws from all parameters (see Table A7.1). To overcome computational challenges with running PSA in a DES for a sample of 100 million women, regression methods were used. A generalised additive model (GAM) was used to predict total costs and total QALYs, for each iteration of the PSA results, using the model parameters as predictors. The fitted values generated from the estimated regression equation then replaced the original total cost and QALY data points.

Table A7.1: Input parameters with sampling distributions and hyperparameters used in PSA

Parameter	Distribution	Hyperparameter 1	Hyperparameter 2
Mammographic Sensitvity		Mean	Standard deviation
Beta 1	Normal	1.47	0.1
Beta 2	Normal	6.51	0.5
VDG modifiers		α	β
VDG 1	Beta	9.538	1.9535
VDG 2	Beta	37.59	8.421
VDG 3	Beta	16.126	7.245
VDG 4	Beta	2.346	1.657
Supplemental imaging		α	β
US	Beta	35.89	11927
MRI	Beta	99.495	19799.5
Growth rate distribution		Mean	Standard deviation
α_1	Normal	1.07	0.09
α_2	Normal	1.31	0.11
Survival post-BC		Mean	Correlated draws
γ NPI 1	Multivariate normal (MVN)	-5.413	See Table A6.1i
γ NPI 2	MVN	-4.023	See Table A6.1i
γ NPI 3	MVN	-2.465	See Table A6.1i
Utility weights			
Early – year 1	1-exp(MVN)	-1.19	See Table A6.1ii
Early – following years	1-exp(MVN)	-1.51	See Table A6.1ii
Advanced cancer	1-exp(MVN)	-1.16	See Table A6.1ii
Costs			
DCIS	(log) MVN	9.08	See Table A6.1iii
NPI 1	(log) MVN	9.3	See Table A6.1iii
NPI 2	(log) MVN	9.47	See Table A6.1iii
NPI 3	(log) MVN	9.64	See Table A6.1iii
Advanced cancer	(log) MVN	10.06	See Table A6.1iii

Table A7.1i: Covariance matrix for survival post-BC

0.015874		
0.008	0.029739	
-0.008	-0.008	0.008

Table A7.1ii: Covariance matrix for utility weights[1]

0.021208		
0.004979	0.012986	
0.007611	0.005956	0.030349

[1]Assumed correlation: 0.3

Table A7.1iii: Covariance matrix for costs[1]

0.5				
0.15	0.5			
0.15	0.15	0.5		
0.157321	0.157321	0.157321	0.55	
0.171026	0.171026	0.171026	0.179374	0.65

[1]Assumed correlation: 0.3

Full incremental cost-effectiveness analysis

A full incremental cost-effectiveness analysis of the base case is presented in Table A7.2. This involves ranking the interventions in increasing order of QALYs and then comparing to the next highest alternative. This shows which alternative are dominated. alternative.

Table A7.2 Full incremental cost-effectiveness analysis

Alternative	**QALYs**	**Costs**	**Incremental cost-effectiveness ratio (ICER); £ per QALY ***
No screening	17.6919	246	
Current UK-NBSP	17.7095	654	23,197
Masking	17.7102	809	dominated by current NBSP
Risk-1	17.7119	694	16,689
Risk-1 and Masking	17.7124	870	dominated by risk-1
Risk-2	17.7181	858	26,749

* base case, discount rate for costs and QALYs 3.5%

Supplementary Appendices References

[1] Evans DG, Astley S, Stavrinos P, Harkness E, Donnelly LS, Dawe S, Jacob I, Harvie M, Cuzick J, Brentnall A, Wilson M, Harrison F, Payne K, & Howell A. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and Family History Clinic: A dual cohort study. NIHR Programme Grant for Applied Research. (NIHRPG-0707-100311). 2016

[2] Baker RD. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Manag Sci 1998;1:103–13.

[3] Tan KHX, Simonella L, Wee HL, Roellin A, Lim Y-W, Lim W-Y, et al. Quantifying the natural history of breast cancer. Br J Cancer 2013;109:2035–43. doi:10.1038/bjc.2013.471.

[4] Van Oortmarssen GJ, Habbema JDF, Van Der Maas PJ, De Koning HJ, Collette HJ a., Verbeek ALM, et al. A model for breast cancer screening. Cancer 1990;66:1601–12. doi:10.1002/1097-0142(19901001)66:7<1601::AID-CNCR2820660727>3.0.CO;2-O.

[5] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Res 2008;10:R41. doi:10.1186/bcr2092.

[6] Lee S, Zelen M. Modelling the early detection of breast cancer. Ann Oncol 2003.

[7] Karnon J, Goyder E, Tappenden P, McPhie S. A review and critique of modelling in prioritising and designing screening programmes. 2007.

[8] Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361–6.

[9] Blamey RW, Hornmark-Stenstam B, Ball G, Blichert-Toft M, Cataliotti L, Fourquet A, et al. ONCOPOOL - a European database for 16,944 cases of breast cancer. Eur J Cancer 2010;46:56–71. doi:10.1016/j.ejca.2009.09.009.

[10] Lundin J, Lehtimäki T, Lundin M, Holli K, Elomaa L, Turpeenniemi-Hujanen T, et al. Generalisability of survival estimates for patients with breast cancer--a comparison across two population-based series. Eur J Cancer 2006;42:3228–35. doi:10.1016/j.ejca.2006.06.028.

[11] Callagy GM, Pharoah PD, Pinder SE, Hsu FD, Nielsen TO, Ragaz J, et al. Bcl-2 Is a Prognostic Marker in Breast Cancer Independently of the Nottingham Prognostic Index Bcl-2 Is a Prognostic Marker in Breast Cancer Independently of the Nottingham Prognostic Index 2006;12:2468–75. doi:10.1158/1078-0432.CCR-05-2719.

[12] Hwang KT, Woo JW, Shin HC, Kim HS, Ahn SK, Moon HG, et al. Prognostic influence of BCL2 expression in breast cancer. Int J Cancer 2012;131:E1109–19. doi:Doi 10.1002/Ijc.27539.

[13] Sobin LH, Gospodarowicz MK (Mary K., Wittekind C (Christian), International Union against Cancer. TNM classification of malignant tumours. Wiley-Blackwell; 2009.