Health Care Management Science 1 (1998) 103–113

Use of a mathematical model to evaluate breast cancer screening policy

Rose D. Baker

Centre for OR and Applied Statistics, Department of Mathematics, University of Salford, Salford M5 4WT, UK

A model of breast cancer screening was developed, in which the processes of tumour origination and growth, detection of tumours at screening, presentation of women with cancers to their GP, and of survival after diagnosis were modelled parametrically. The model was fitted to data from the North-West of the UK, for 413 women who screened positive, and for 761 women who developed interval cancers. Model validation comprised verification that the final model fitted the data adequately, together with the comparison of model predictions with findings by other workers.

The mathematical model was used to assess different screening policies, and to ask "what if" questions. Taking the cost of breast cancer to be the sum of the cost of screening and the cost of PYLL (person years of life lost due to cancer), the optimal screening policy was calculated. The costs of the current policy and of other possible screening policies were found, together with their effects on life lost and on mortality. The tentative conclusion was that if monies can be found to extend the screening programme, for example to carry out one more screen per woman, most benefit would be obtained by reducing the start age of screening by 3 years.

1. Introduction

1.1. The medical screening problem

In the medical screening problem, we wish to evaluate the cost and benefits of screening policies in which screening of an individual takes place at a "vector" of ages **a**, i.e., $a_1, \ldots, a_N$. For breast cancer, screening started in the North-West (UK) in 1988, and the current policy is to screen women every 3 years from age 50 to age 64. There is currently debate as to whether the screening interval should be reduced to 2 years, or whether the screening age-band should be extended up into the late 60's and early 70's, or down into the 40–50 age band.

The rationale of screening is that the prognosis is better when treating the smaller tumours that can then be detected. The aim is to choose a screening policy to minimise the total cost of cancer. We therefore seek to minimise the cost of cancer plus the cost of carrying out any screening.

1.2. Mathematical modelling

Mathematical modelling of some sort is necessary for at least two reasons. One is that modelling enables us to assess the benefits of screening policies without carrying out very extensive and time-consuming controlled trials, which may also be ethically dubious. Modelling enables this benefit to be estimated using data that are already available.

Another reason is that even with a controlled trial, we can only assess the screening policy used in that trial; without modelling we cannot assess other possible policies, or find the optimum (minimum total cost) policy.

Interest in modelling cancer screening has therefore been great. Duffy et al. [2] reference much cancer screening modelling work. Simpler models of the mathematical type proposed here were developed by several workers some years ago [1,9], but a simulation approach based on a Markov approximation (the MISCAN program, e.g., Oortmarssen et al. [6]) is today widely used to evaluate screening policies.

1.3. How this model differs from others

The problems a modeller must face are those of: choosing a suitable model; fitting it to data to estimate values of the model parameters; validating and refining the model; and finally, using it to evaluate different possible screening policies.

In the MISCAN approach, state variables, such as tumour size, are lumped into several classes, and a woman moves from class to class with exponential dwell times in each class (i.e., this is a Markov model). The model is fitted to data and validated by simulating many life-histories from the model, and by tweaking model parameters until the various model predictions agree with summary statistics drawn from observed data. Assessment of screening policies is also carried out by simulating many life-histories with a given policy in force, and so estimating the cost.

The present work differs from the MISCAN approach in many ways. State variables (such as tumour size) are not "polychotomised" into classes, but are retained as continuous variables that change smoothly with time. The Markov assumption gives mathematical tractability, but at the expense of realism, and so is not made here.

Model fitting is done using the standard statistical technique of "maximum likelihood estimation". This requires the calculation of the *likelihood function*, which is a complicated function of data and model parameters that for this problem requires numerical evaluation of double and triple

integrals. The drawback over the MISCAN method is the difficulty of doing the mathematical calculations and programming them onto a computer. The advantage is that this method of estimating model parameters is more efficient, in that all the information in the data is used, giving smaller errors on fitted model parameters. Another advantage is that it is now possible to decide easily which of two possible model parameterisations fits the data better. The model fitting and validation procedure is now based on standard statistical theory, and so is more automatic and less *ad hoc*.

Finally, assessment of screening policies is done by computing the cost of a policy as a complex function of model parameters, to be calculated directly, rather than through simulation of many case-histories from the model.

Model fitting and policy evaluation was done on a Pentium 120 PC using Fortran95 programs written by the author, and NAG routines (see, e.g., [3]) were used for function minimisation, Gaussian integration, etc.

The following sections describe in more detail the screening model, its fitting to data, and the evaluation of various policies, including the one currently adopted.

2. The screening model

Model assumptions are as follows:

(1) tumours "originate" at a rate that increases with age. The lifetime risk of breast cancer is only 8%, so that multiple tumour originations are rare and the process is modelled with a simple pdf of tumour origination;

(2) tumours once originated grow almost exponentially with time. The rate of growth slows down exponentially with age;

(3) there is a distribution of tumour growth-rate;

(4) the sensitivity of tumour detection at screening is an increasing function of tumour size, and was allowed to vary with age (no such variation was however observed from fitting data from women aged 50–65);

(5) probabilities of detecting tumours at successive screenings are independent. Nothing is known empirically about the validity of this assumption, which is probably not quite correct, because some tumours may be harder to detect than others throughout their growth, due to location, shape, or degree of calcification;

(6) the probability that a woman with a cancerous tumour of given size decides to present to her GP is a function only of tumour size and tumour growth-rate, but not of time from last screening;

(7) the hazard of death after diagnosis is a function of tumour size and tumour growth-rate, and initially increases after diagnosis, before decreasing to near the normal age-dependent hazard for women without cancer;

(8) screening has no harmful effect. The number of cancers caused by the X-rays used in mammography is believed to be very small and difficult to quantify (Jansen and Zoetelief [4]);

(9) DCIS (ductal carcinoma *in situ*) tumours seldom metastasize, and are hence not very harmful. They are currently ignored in the model;

(10) the cost function $\mathcal{C}$ is taken as the cost of the expected number of months of life lost due to cancer per woman, plus the cost of the screening programme.

All parameterisations were *ad hoc*, to best fit the data, except for the near-exponential growth of small tumours, which is well known.

The following sections describe the model in more detail.

2.1. Tumour origination and growth

The probability density function (pdf) $f_a(a)$ of a tumour originating at age a is required. Since this probability is small, it is approximately equal to the hazard function $h(a)$.

This was modelled as

$$h(a) = \frac{1}{a_1 \exp(-\theta_1 a) + a_2 \exp(-\theta_2 a)},$$

a form chosen as likely to be able to reproduce the shape of the pdf of presentation as a function of age in the absence of screening. This latter curve (figure 1) must be closely related to the pdf for tumour origination, as presentation follows origination after a variable but small time lag, of not more than a few years.

Tumours are assumed to originate at size (diameter) ω_0. It would be more biologically realistic to insist that ω_0 was the size of 1 cell, but allowing ω_0 to be fitted to data gives more freedom in the parameterisation of screen sensitivity.

The simplest growth assumption is that of exponential growth. This is reasonable for the comparatively small tumours which are detected at screening, or with which

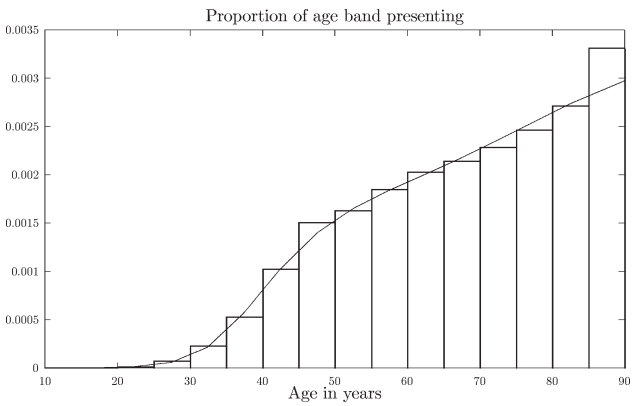

Figure 1. Observed and predicted proportions of women in successive age-bands presenting with cancer.

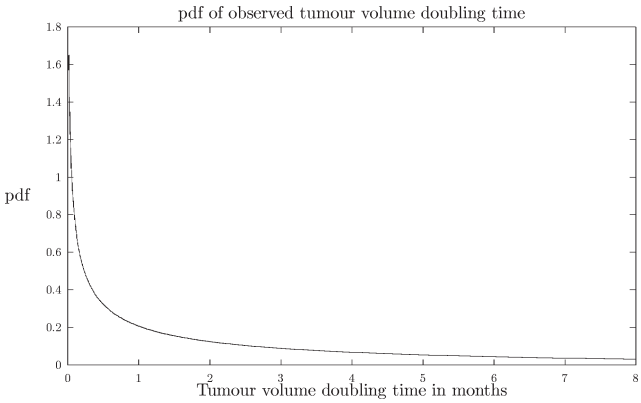

Figure 2. Frequency of tumour doubling times for women in the 50–65 age-range, predicted from the model.

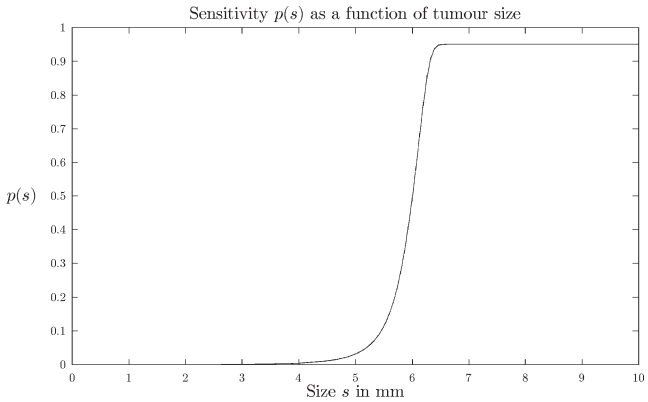

Figure 3. Sensitivity of screening as a function of tumour size, as predicted by the model.

women first present. If tumour diameter s grows exponentially, this gives the differential equation

$$\mathrm{d}s/\mathrm{d}t = rs,$$

where t is time from origination.

However, the rate of tumour growth slows with age, so that tumours grow progressively more slowly during the course of the disease. This can be parameterised as

$$\mathrm{d}s/\mathrm{d}t = rs\exp(-\phi t)$$

at age t, and $\phi \geqslant 0$. The solution is the Gompertz curve

$$(\phi/r)\log(s/\omega_0) = \exp(-\phi a_0) - \exp(-\phi t), \qquad (1)$$

where the tumour size is ω_0 at its origination at age a_0.

It will be necessary to turn equation (1) round, to give the age at origination of a tumour in terms of its size at age a. The solution is

$$a_0 = a - \frac{\log(1 + (\phi/r)\exp(\phi a)\log(s/\omega_0))}{\phi}. \qquad (2)$$

The rate of tumour growth r is assumed to vary randomly from person to person, so that r has pdf $f_r(r)$. The distribution of r or r^{-1} could be modelled using any survival distribution. The gamma, Weibull, lognormal and Pareto distributions were fitted, but the best fit was obtained with a Weibull distribution for f_r. Figure 2 shows the fitted distribution of tumour doubling times.

2.2. Screening

At screening, a tumour of size s will be detected with probability (sensitivity) $p(s, a)$ in a woman of age a. It is assumed that successive screenings constitute independent Bernoulli trials. The sensitivity must satisfy $p(0, a) = 0$, $p(\infty, a) = \upsilon$, where υ is the sensitivity towards a very large tumour. An obvious choice is $\upsilon = 1$, but it was constrained to be 0.95, from knowledge about the screening process. The best-fitting choice for $p(s, a)$ to date is

$$p(s, a) = \upsilon\big\{1 - \exp\big(-((\exp(\kappa s) - 1)/\eta)^{\gamma}\big)\big\},$$

where $s > \omega_0$; otherwise $p(s, a) = 0$.

In the 50–65 age-range considered, screening sensitivity was found not to be a function of age.

Figure 3 shows the fitted sensitivity as a function of tumour diameter.

2.3. Presentation

The probability Q that a woman has presented to her GP must surely increase with tumour size. It may also be a function of tumour growth-rate [8]. Since tumour size at origination ω_0 is a constant, Q can be written as $Q = Q(\log(s/\omega_0))$, where $\log s/\omega_0 \propto \log_2(s/\omega_0)$, the number of doublings since tumour origination. Including tumour growth-rate r to give $Q(\log(s/\omega_0) + cr)$ did not improve the model fit.

The best fitting model for Q was the normal distribution function, $Q(\log(s/\omega_0)) = \Phi(\log(s/\omega_0); \mu, \sigma^2)$, where μ is the mean and σ^2 the variance. The pdf corresponding to $Q(x)$ is denoted as $q(x)$.

2.4. Cancer-free survival

Mortality from all causes results in a probability $R(t)$ that a woman still lives at age t (R is the survival function). The best functional form for R was found to be the Gompertz distribution, with exponentially increasing hazard of death $h(t) = \alpha\exp(\beta t)$ at age t. A current life table was available, of number of deaths n_i in a period of $\Delta t = 1$ year, among N_i women in the ith age band (a_{i-1}, a_i). The model was fitted by minimising the chi-squared

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - N_i P_i)^2}{N_i P_i (1 - P_i)},$$

where P_i is the model prediction.

Table 1 shows fitted model parameter values. The chi-squared was 98 with 14 degrees of freedom. This would not be regarded statistically as a good fit; however the curve approximates closely to the data, being a little less steep than required at low ages, and is adequate for practical purposes. The large size of the data sample enables very small model discrepancies to become significant.

Table 1
Model parameter values for fit to 1991 death-rate table.

Model parameter	Meaning	Value
α	Scale	0.242710^{-4}/year
β	Exponent multiplier	0.09809/year

2.5. Survival after diagnosis

The survival data was first fitted to a survival model with tumour size at diagnosis as a covariate. Later, tumour growth-rate r was added as a covariate that could affect the hazard of death, and as this is not directly measured, it was then necessary to estimate survival parameters by maximising the full likelihood function, which includes the modelling of screening and presentation.

It is known that after diagnosis the hazard of death from cancer initially increases, as the tumour must grow and metastasize to cause death. However, some tumours can be dealt with, and if death has not occurred within a few years, mortality is roughly that appropriate to a cancer-free woman of the same age. The hazard of death therefore initially increases and subsequently decreases.

Let p be the probability that a tumour metastasizes and would certainly lead to death, S_b and h_b be the survival function and hazard of death from a metastasized tumour, and $S_0(t)$ the survival function for a period t from other causes of death, given survival to age of diagnosis.

Then the total survival function is

$$S(t) = S_0(t)\{pS_b(t) + 1 - p\},$$

and the hazard

$$h(t) = h_0(t) + ph_b(t)S_b(t)/(pS_b(t) + 1 - p),$$

where h_0 is the hazard pertaining to the survival function S_0. This behaves as required, if h_b increases with t.

The best-fitting model was found with S as an IFOM (increasing force of mortality, i.e., hazard) Weibull survival function, and p as an increasing function of tumour size s, $p = 1/(1 + \mathrm{e}^{\xi} s^{\varepsilon})$. On including tumour growth-rate r, the best-fitting Weibull survival function was

$$S_b(t) = \exp\bigl(-\bigl(\lambda(r\exp(-a\phi))^{s^{\delta}} t\bigr)^{\nu}\bigr).$$

Here the probability that a tumour leads to death increases with tumour size, but the time to death from it is a function of an interaction term between growth-rate and tumour size. More data would be needed to refine this parameterisation.

With the Gompertz distribution for mortality from other causes, S_0 becomes

$$S_0(a,t) = \exp\bigl\{-(\alpha/\beta)\bigl(\exp(\beta(t+a)) - \exp(\beta a)\bigr)\bigr\},$$

where a is age at diagnosis.

Figures 4 and 5 show observed and predicted numbers of deaths following screening and presentation respectively. The model reproduces the worse prognosis after presentation. The poor fit in the tail of figure 5 is probably due to women who presented late in calendar time, and for whom the follow-up period was short. This censoring was taken into account in the maximum-likelihood fit of the model to data, but is currently not adjusted for in the plotted histogram.

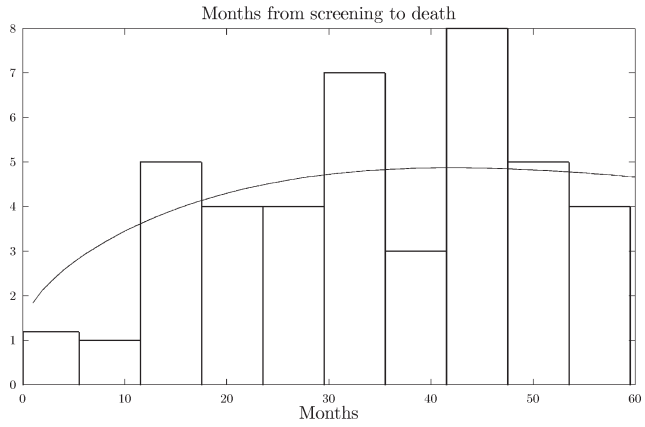

Figure 4. Observed and predicted deaths by month from screening.

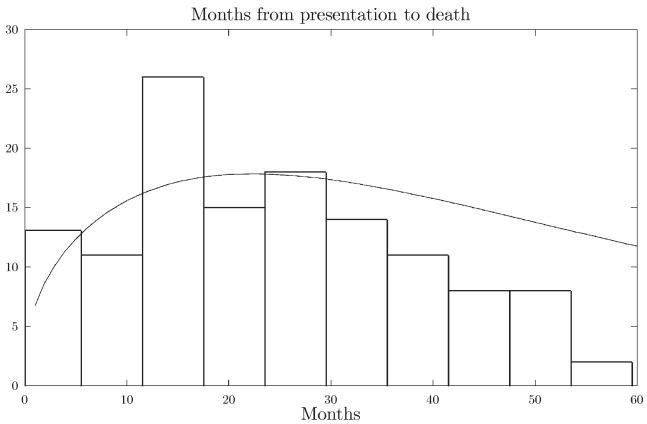

Figure 5. Observed and predicted deaths by month from presentation.

3. Parameter estimation

3.1. Outline of procedure

Five types of data were used in this analysis:

(1) tumour size for cancers detected at screening. This comprised 761 cases from screens taken between 1988 and 1990 in screening centres at Manchester, Wigan and Bolton;

(2) timing and tumour size for interval cancers. This comprised 413 interval cancers occurring between 1989 and 1995 at the above three screening centres;

(3) survival periods of women who had tumours found on screening or at presentation, after cancer had been diagnosed. This comprised 716+431 cases, with the end of follow-up at 31 December, 1996;

(4) a table of 1987 cancer incidence in the NW by age band for the period before screening was instituted, comprising 39170 cases;

(5) a 1991 survival table for death from all causes for women in the North West, containing 21678 deaths, created by CCE.

The data were from Manchester, Wigan and Bolton, both for women whose tumours were detected at screening, and for those presenting with cancer. Tumour size was sometimes missing, as was age. These data concerned only women who had cancer, and gave a complete picture of the progress of the disease, from detection to death. They had to be supplemented by pre-screening cancer incidence data, to enable the incidence rate to be estimated, and by a survival table, as the survival function of death from other causes enters the model.

The likelihood function for the screening and interval cancer datasets is the likelihood $\mathcal{L}_c$ of observing cancer surfacing in the way it did (at screening or on presentation), given that the woman would have been present in the cancer database.

The contribution to the conditional likelihood to be maximised from one woman is then

$$\mathcal{L}_c = \mathcal{L}/\int \mathcal{L},$$

where the integral in the denominator runs over all variables such as tumour size, and is the probability that the woman appears in the database.

To obtain a likelihood function for all the data available, the product of conditional likelihoods taken over all women in the database is multiplied by the likelihood of observing the pre-screening cancer incidence data. The model was fitted by choosing model parameters to maximise this total likelihood. For numerical convenience the log-likelihood was computed and maximised rather than the likelihood.

For fitting the life-table, and for fitting the contribution to the log-likelihood function due to pre-screening cancer incidence data, the log-likelihood was replaced by minus half of a chi-squared. The chi-squared is simply a large sample approximation to $\chi^2 = -2(\ell - \ell_0)$, where ℓ is the log-likelihood and ℓ_0 the log-likelihood of the saturated model. Using the chi-squared enables goodness of fit to the life-table data to be assessed.

The computation of the likelihood function is mathematically technical and is described in appendix A.

4. Model validation

Model parameters fitted by maximising the conditional likelihood are described in table 2.

Models were refined by trying several different parameterisations, and choosing the one that minimised the Akaike Information Criterion (AIC). The minimum-AIC estimator (MAICE) model was taken as giving the best fit to the data.

Because procedures for maximisation of nonlinear functions can converge to false (local) maxima, a number of trials were carried out in which parameter values were randomly varied from their optimum values, and the likelihood re-maximised. This gave confidence that the global maximum had been found.

Goodness of model fit was assessed visually by plotting data and model predictions for a variety of marginal distributions, such as tumour size, time of presentation from last

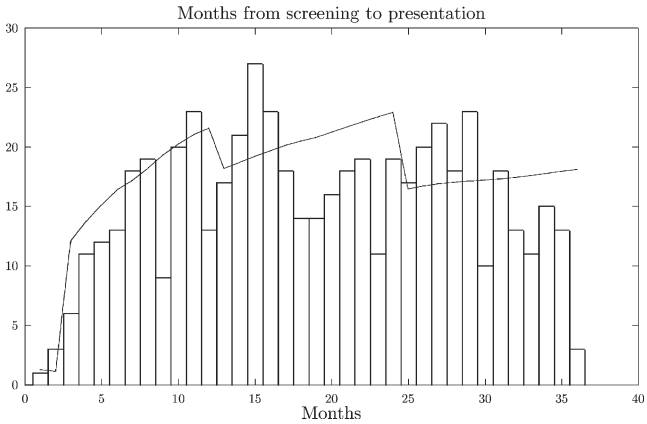

Figure 6. Observed and predicted frequencies of months from last screening to presentation with cancer.

Table 2
Model parameter values for fit to screening and interval cancer data and to cancer incidence table.

Model parameter	Meaning	Value
a_1	Tumour origination scale parameter	0.689×10^7 months
θ_1	Tumour origination rate parameter	0.026/month
a_2	Tumour origination scale parameter	183.2 months
θ_2	Tumour origination rate parameter	0.00124/month
ω_0	Initial tumour size	0.250 mm
ψ	Tumour growth scale	0.456×10^8 months
χ	Tumour growth shape	0.0978
ϕ	Tumour growth slowdown	0.0017/month
κ	Screening sensitivity parameter	275/mm
η	Screening sensitivity parameter	0.250
γ	Screening sensitivity parameter	4.51
υ	Screening sensitivity parameter	0.95 (fixed)
μ	Presentation distribution mean	4.54
σ	Presentation distribution sdev.	0.535

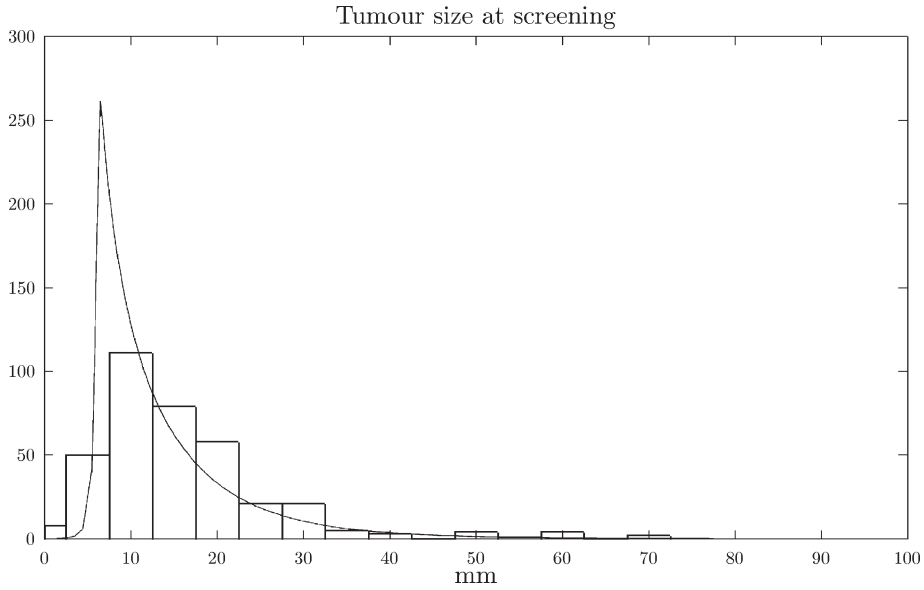

Figure 7. Observed and predicted tumour sizes at screening.

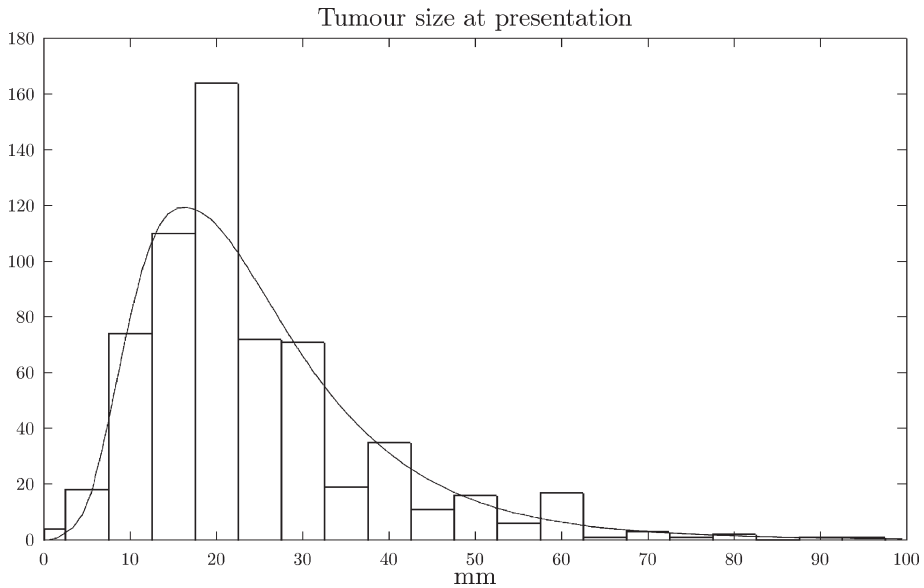

Figure 8. Observed and predicted tumour sizes at presentation.

screening, and so on. Figures 1 and 6–8 are a sample of about 20 plots of this type that were used. The overprediction of interval cancers in the first 3 months after screening is thought to be due to the model's neglect of psychology. After being screened for cancer, women are less likely to attribute pain or breast lumps to a tumour. They may also be waiting for screen results. The jumps in the model prediction in this plot are due to the removal of women who presented in years of incomplete follow-up from the data.

A variety of model extensions were explored, to obtain the best-fitting model. These methods ensure "internal validity", that the model fits the data.

Another method of model validation to ensure "external validity" is to compare model predictions with other knowledge besides the data fitted. Here such predictions include screening test sensitivity and the distribution of tumour growth-rates at presentation.

The maximum likelihood estimate of sensitivity would tend to unity as tumour size increases. To ensure compatibility with other knowledge, it has been constrained not to exceed 0.95. The average sensitivity for tumours of less than 10 mm is then 0.61 for tumours in the data sample. This compares with the value of 0.7 cited by Oortmarssen et al. [6]. The sensitivity has already attained its value of 0.95 for tumours of 10–19 mm, in agreement with Oortmarssen et al.

The distribution of tumour growth-rates has median tumour volume doubling time of 88 days, mean volume doubling time of 175 days (5.7 months). This is consistent with many studies cited by Johnson and Shekdar [5].

Table 3
Model parameter values for fit to survival data after diagnosis, to parameterise $S_b(t) = \exp(-(\lambda(r\exp(-a\phi))^{s^\delta}t)^\nu)$ and $p = 1/(1+\mathrm{e}^{\xi}s^{\varepsilon})$.

Model parameter	Meaning	Value
λ	Weibull scale parameter for S_b	0.01696
ν	Weibull shape parameter for S_b	1.626
δ	s coeff. in power of tumour growth-rate in S_b	1.448
ε	Tumour size power in probability p	−2.113
ξ	Constant term in p	7.32

The slowdown of tumour growth with age predicted by the model is approximately in agreement with data from Peer, as cited by Jansen and Zoetelief [4]. Peer finds a ratio of 2.3 of tumour volume doubling times between age groups of average ages 45 and 77 years. The model presented here predicts a ratio of $\exp(\phi\times 32) = \exp(\times 0.0017\times 12\times 32) = 1.92$.

5. Evaluation of screening policies

5.1. Calculation of optimum policy

Given the model and parameter estimates, it is possible to calculate any desired loss function (cost), conditional on a vector **a** of screening ages. It is thus possible to choose the screening policy (vector of screening ages) that minimises the loss function. The loss function used in this work is the total cost of person years of life lost due to cancer (PYLL) and of screening.

For a growth-rate r, the expected number of years of life lost due to cancer after diagnosis at age a is

$$L(s,r,a) = p\int_0^\infty S_0(1-S_b)\,\mathrm{d}t.$$

The loss function is then

$$\begin{aligned}\mathcal{C}' = C_y\sum_{i=1}^{N}\int_0^\infty\int_0^\infty L(s,r,a_i)f(s,r,\mathbf{a}_i)f_r(r)\,\mathrm{d}r\,\mathrm{d}s\\ +\sum_{i=1}^{N+1}\int_{a_{i-1}}^{a_i}\int_0^\infty L(s,r,t)g(s_p,r,t\mid\mathbf{a}_{i-1})f_r(r)\,\mathrm{d}t\,\mathrm{d}s_p\\ +C_s\sum_{i=1}^{N}R(a_i),\end{aligned}\tag{3}$$

where C_s is the cost of 1 screen, and C_y is the cost of one month of life. Define $\mathcal{C} = \mathcal{C}'/C_y$. Minimisation of $\mathcal{C}'$ is equivalent to minimisation of $\mathcal{C}$, which is a function only of the ratio of costs C_y/C_s. To evaluate the cost of a policy therefore requires the specification of the number of screens that could be paid for with the cost of one month of life.

For small numbers of screens per woman, it was possible by discretizing age to the nearest year to explore all possibilities, and so to find the optimum policy.

For more than 3 screens per woman, it was much faster to use the NAG function minimiser E04JBF to find the optimum policy. As before, iteration was restarted from random screening ages to ensure that the global minimum had been found.

On setting the probability of detecting a tumour to zero for any screens earlier than the most recent, it became possible to calculate the optimum policy using dynamic programming. The cost function given any policy is a sum of terms, each term corresponding to the path between two successive screens. These path costs now depend only on the path itself (the start and finish screen ages) and not on previous events. This is because the functions $f(s\mid\mathbf{a})$ and $g(s_p,t\mid\mathbf{a})$ are now functions only of the most recent screening age. This Markov assumption that earlier screens have negligible effect is good unless screens are much closer together than would ever be contemplated. Paths were in fact taken between ages 6 months apart.

The algorithm programmed works by building up the cheapest path from the start to the end age points (n time units apart). All possible segments of length 1 are found, then all cheapest segments of length 2 are built from these, and so on until the cheapest path from the start age to the finish age is built up. The computation of the cheapest path was very fast given the $n(n-1)/2$ costs of paths between pairs of ages without intervening screens.

5.2. The impact of various screening policies

Table 4 shows the costs of various policies, including the optimal policy found using dynamic programming.

A range of optimal policies can be derived, by varying the value we place on human life. Assuming that 8 screens are equal in cost to one month of life, the model predicts that 6 screens are optimal. As a screen costs about £25, this places a value of £2400 on a year of human life. This is thus roughly the implied value of human life from the existing policy, in which women are screened 3-yearly from age 49/50 to age 64, so that the aim is to give 6 screens in all.

The results confirm that the benefit of screening is very modest, with even annual screening over a wide age-range (48–64) producing only a 26% reduction in years of life lost due to cancer. The optimum policy requires screens roughly every 2.5–3 years from age 48 to age 61.5. Intuitively, this is because in the 50's the risk of developing a tumour has become high, and the tumour growth-rate is still high, so that frequent screening is desirable, while the benefit in terms of years of life to be saved is also still high. In older women, tumours grow more slowly and the number of years of life that can be salvaged is lower. In younger women, tumours are less likely to arise.

Doubling the cost of months of life lost, 16 screens are equivalent to 1 month of life, and the optimum ages are now: 40.5, 42.5, 44.5, 46.5, 48.5, 51, 53.5, 56, 59, 62, 65, 68.

This is essentially a pattern of screens every 2 years from age 40–50, with screens every 3 years from age 50 to age 70. It is curious that as the cost of screening decreases,

Table 4
Screening policies evaluated under the cost function in equation (3). % R_1 is the percentage reduction in years of life lost due to cancer, %R_2 is the percentage reduction in numbers of deaths from cancer. Cost is measured in months of life, and includes screening cost, assuming 1 month will buy 8 screens. DP is "dynamic programming". The 3rd entry is the current policy, plus one additional screen, chosen to minimise cost.

Policy	Screening ages	Cost $\mathcal{C}$	% R_1	% R_2
None		5.619	0	0
Current	51.5–63.5	5.5704	11.2	13.0
2-view	51.5–63.5	5.48	12.8	14.7
Current + lower	48, 51.5–63.5	5.56	13.5	14.9
Biannual	48–64	5.908	17.8	19.0
Annual	48–64	6.1822	25.4	25.8
Optimum	48, 50.5, 53, 55.5, 58.5, 61.5	5.487	15.0	15.25
Annual	48–70	6.557	29.8	33.4

the optimum policy requires screening most intensively in the 40–48 age-range, in which at higher screening cost, no screens at all would have been done. This is consistent with repeated claims made in the cancer literature that more frequent screening is required in the 40–50 age band, precisely because the benefit of screening there is smaller.

The logic of the model is analogous to that of a man who has limited time in which to pick up a heap of dropped money. He will first pick up the large coins. With more time at his disposal, he will go on to pick up the smaller coins. It could be that in that case most of his time is in fact spent picking up small coins, even although such activity contributes little to the total he picks up.

Similarly, the model recommends most screening effort going where it reduces PYLL due to cancer only slightly, given sufficient money devoted to screening.

It is also useful to specify the next feasible step towards optimality from the current situation. Screening is in fact carried out as follows: screening teams make a circular tour of all GP practices, and women who are of age at least 49/50 are invited to attend. Screening will not be available in an area again for another 3 years. This means that there is little point in specifying optimum screening ages, because any individual woman can only attend when the team is in her area. If just under 50 at a screening round, she will be nearly 53 at first screen.

Progress towards an optimum policy would start by retaining the 3-year cycle, but finding extra workers so that women outside the 50–64 age-band could be screened. The question of immediate interest is then: if we can afford to do one extra screen, should we extend the age-range upwards or downwards?

To answer this, the cost programme was run with screening ages 51.5–63.5, at 3-year intervals, with the addition of one extra screen. The ages 51.5, 54.5, 57.7, 60.5, 63.5 were chosen to be average ages at 1st to 5th screens. As table 4 shows, to minimise cost the extra screen should be at age 48, 3.5 years *below* the age at first screen.

The data described here were obtained from 1-view mammography. It is possible to move to 2-view mammography, in which the sensitivity of detecting a tumour is greater. Screening cost is higher, and it is of interest to estimate the gain. Will this move be cost-effective, and greatly reduce years of life lost due to breast cancer?

Sadly, the answer to this question from the modelling work here is "no". Assuming certain detection of all tumours greater than 5 mm in size, table 4 shows the percentage reduction in mortality and years of life lost by continuing with the current screening policy, but at this heightened sensitivity. Gains in reduction of percentage mortality and years of life lost are only 1.6–1.7%, an extremely modest improvement over the present situation.

The reason undoubtedly lies in the wide distribution of tumour growth-rate. Many tumours that were less than 5 mm in size at screening will have grown dramatically, and the affected woman will have presented with symptoms of cancer before her next screen 3 years later.

6. How reliable are the model predictions?

In general, assuming correct mathematical development and computer programming, model predictions may be wrong because:

(1) the sample of data is small, giving large statistical errors on fitted model parameters;

(2) the model's parameterisation does not allow it to fit the data well;

(3) the model ignores biases in the data;

(4) the model is used to extrapolate (for example, in age) away from the age-range of the data to which it was fitted.

In this study, the data sample was large enough for statistical error on fitted model parameters to be small, and the parameterisation was varied so that the fit to data appeared reasonable.

The effects of bias introduced because women choosing to be screened may be at higher risk of cancer than others is complex. The conclusion arrived at here is that the model as currently fitted is correct, if applied to the whole population of women, or to a random woman from the general population.

The sample of women who were screened, some of whom then developed interval cancers, were all women who accepted the invitation to attend for screening. They might well have done so because they perceived themselves as being at high risk of developing breast cancer. However, if the hazard of developing cancer at age a for such women is simply proportional to the hazard obtaining for the general population, this will not affect the fitted model parameters. This is because such a scaling factor would cancel out from the conditional likelihood $\mathcal{L}_c$. The scale of the hazard function is determined by the fit to cancer incidence data obtained before screening began.

Thus although the model was fitted to data from women who chose to be screened, its predictions are for the general population. They are not biased, given 100% compliance. If it were desired to assess the impact of screening only the self-selected group who choose to be screened, it would be necessary to estimate their hazard of cancer relative to the general population. This could be attempted, given data on cancer rates in comparable women who did and did not choose to be screened.

The model was fitted to data from women in the 50–65 age-group. In this age-range, the sensitivity of screening did not increase with age. It is however thought that tumours are harder to detect in younger women.

The model prediction is that the UK screening policy would improve if women were given an extra screen 3 years before the current first screen around age 50. This is a strong conclusion. However, such conclusions about the benefits of screening down to age 40 must be regarded as tentative until data on the effect of age on screening sensitivity can be used to refine the model.

7. Conclusions

A statistical/mathematical modelling/OR approach to evaluating screening policies has been presented. Despite nontrivial numerical and statistical problems, the methodology delivers useful results.

Compared with a simulation approach such as that used by MISCAN, the maximum likelihood estimation of model parameters has some advantages and disadvantages. Advantages are that the information in available data is used as fully as possible in the estimation of model parameters. It is also possible to choose the best of several alternative parameterisations as that with the minimum AIC.

It is however true that the process of model-fitting requires effort to solve numerical problems, such as the evaluation of multiple integrals and the ability to reach the global maximum of the likelihood function. The process of deducing and rectifying model inadequacies from plots of model predictions against data is also a difficult one, requiring mathematical insight.

The main areas remaining to be addressed are sensitivity analysis and the modelling of non-compliance.

Regarding sensitivity, some model parameters may be poorly determined, but may also have little impact on policy evaluation. We need to know which model parameters are both poorly-determined, and also strongly affect model predictions on optimum policy.

Incomplete compliance with the screening programme will affect the optimum policy if modelled as a function of age, unless women can be dichotomised as "always respond" and "never respond" to the request to attend for screening. Only about 70% of women attend for screening, and compliance is thought to fall off with age, which would bias the optimum policy further towards younger women.

Acknowledgement

I would like to thank Professor Cieran Woodman and Anthony Threlfall of the Centre for Cancer Epidemiology, Manchester, for extensive help and advice.

Appendix A

A.1. The likelihood function

As women behave independently, the likelihood of the observed data (tumour sizes and timing of presentation to GP) is the product of likelihood functions for each individual woman in the sample. We therefore focus on this individual likelihood. Let **a** be the N-fold vector of ages at successive screens, $a_1, \ldots, a_N$. It is useful to refer to the ith age as a_i, and the vector of the first i ages as $\mathbf{a}_i$. The vector of screens performed by age t is written $\mathbf{a}(t)$.

The timing of screenings is assumed fixed, or at least to be unrelated to the woman's condition, so that the pdf $f_s(\mathbf{a})$ of screens being done at ages $a_1, \ldots, a_N$ is a constant that can be ignored. The likelihood is therefore taken as conditional on age at screening. This assumption would be incorrect if for example women only attended for screening if they had reason to think they might have a tumour.

Let $f(s \mid \mathbf{a}_m)$ denote the pdf of detecting a tumour of size s at the mth of the N screens, and $g(s_p, t \mid \mathbf{a})$ the pdf of a woman presenting with a tumour of size s_p at age t. The functions f and g will shortly be found in terms of model parameters and distributions already introduced. The discussion of the modification of the likelihood function for the addition of data on time to death after diagnosis is postponed for clarity of exposition.

The probability of a tumour being detected in some way (at screening or at presentation) is then

$$\mathcal{P}(\mathbf{a}) = \sum_{m=1}^{N} \int_0^\infty f(s \mid \mathbf{a_m})\,\mathrm{d}s + \int_0^\infty \int_{a_l}^{a_u} g\big(s_p, t \mid \mathbf{a}(t)\big)\,\mathrm{d}t\,\mathrm{d}s_p, \qquad (4)$$

where a_l, a_u are lower and upper age limits of screening (50 and 65).

If a tumour of size s is detected at screening, the conditional likelihood of this, given that the woman appears in

the database, is $f(s \mid \mathbf{a})/\mathcal{P}(\mathbf{a})$. If the tumour size is unknown, the factor contributed to the total conditional likelihood is $\int_{\omega_0}^{\infty} f(s \mid \mathbf{a})\mathrm{d}s/\mathcal{P}(\mathbf{a})$. If the woman presents with the tumour, the factor contributed is $g(s_p, t \mid \mathbf{a})/\mathcal{P}(\mathbf{a})$, and if the tumour size is unknown, the factor is $\int_{\omega_0}^{\infty} g(s_p, t \mid \mathbf{a})\,\mathrm{d}s_p/\mathcal{P}(\mathbf{a})$.

The probability $\mathcal{P}$ of detecting a tumour in some way from equation (4) does not give the true probability of a case appearing in the database, because of slow follow-up: some data have not yet come to hand. Follow-up is complete to end 1994. This has two implications for the computations: interval cancer data in years for which follow-up is not complete cannot be used, and the likelihood denominator $\mathcal{P}$ for other data must be adjusted so that there is no predicted probability of observing such interval cancers.

Further, complete screening results are not available after 1990. Hence when the screening year is later than 1990, the first (f) integral is omitted from $\mathcal{P}$, because the case would not appear in the data file if the cancer had been detected at screening.

In more detail, $\mathcal{P}$ is adjusted by reducing a_u to that maximum age a'_u for a particular individual, such that any interval cancer would occur in a calendar year for which follow-up was complete.

Although interval cancer data in years after screening for which follow-up is incomplete can not contribute factors to the likelihood function in the way described above, it is possible to glean a factor of $g(s_p, t \mid \mathbf{a})/\int_{\omega_0}^{\infty} g(s_p, t \mid \mathbf{a})\,\mathrm{d}s_p$, the pdf for observing a tumour of size s_p when a tumour of any size could have been observed at that time point. Correspondingly, extra screening data from Wigan taken in 1993 contributes a term $f(s \mid \mathbf{a})/\int_{\omega_0}^{\infty} f(s \mid \mathbf{a})\,\mathrm{d}s$.

The probability P_i of presentation in the ith age band (a_{i-1}, a_i) given no previous presentation and no screens is given by

$$P_i = (1/R(a_{i-1}))\int_{a_{i-1}}^{a_i} \mathrm{d}t \int_0^{\infty} g(s_p, t \mid \mathbf{0})\,\mathrm{d}s_p, \qquad (5)$$

where the vector of screening ages $\mathbf{a} = \mathbf{0}$.

If the observed number of presentations is n_i, and the number of women in the age-band is N_i, if the model is correct, for large samples

$$\chi^2 = \sum_{i=1}^{p} \frac{(n_i - N_i P_i)^2}{N_i P_i (1 - P_i)}$$

will obey the chi-squared distribution. Table 2 shows the values of fitted parameters. Data from ages 25–85 were used for fitting. The chi-squared was 73 with 12 degrees of freedom. We can therefore reject the null hypothesis that the model fits the data, but the fit is very good by eye (figure 1); we are therefore willing to use the fitted model. The large sample size has enabled us to detect very small departures of the data from the fitted model.

A.2. The pdf of detecting a tumour of size s at the Nth screening

Consider a tumour that grows at rate r. To be size s at age a_N when screening takes place, it must have originated at time $a_0 = a_N - \log(1 + (\phi/r)\log(s/\omega_0))/\phi$. The pdf of this occurring and so giving a tumour of size $(s, s + \mathrm{d}s)$ is $f_a(a_0)J$, where $J = |\mathrm{d}a_0/\mathrm{d}s|$ is the Jacobian to convert from time interval $\mathrm{d}a_0$ to tumour size $\mathrm{d}s$. From equation (1),

$$J = |\mathrm{d}a_0/\mathrm{d}s| = \frac{1}{sr(1 + (\phi/r)\exp(\phi a_N)\log(s/\omega_0))}.$$

The woman must also not have died by age a_N, with probability $R(a_N)$, or have presented, with probability $Q(\log(s/\omega_0))$. Further, the tumour must not have been detected at $N - 1$ previous screenings, with probability $\prod_{k=1}^{N-1}(1 - p(s_k, a_k))$, and must be detected at the Nth screening, with probability $p(s, a_N)$. Here s_k is tumour size at age a_k, which follows from equation (1) as

$$s_k = s\exp\{-(r/\phi)(\exp(-\phi a_k) - \exp(-\phi a_N))\}.$$

Finally, it is necessary to integrate the pdf over all growth-rates r. This logic gives:

$$\begin{aligned} f(s, \mathbf{a}) &= R(a_N)Q(\log(s/\omega_0)) \\ &\times \int_0^{\infty} \Bigg[f_a\big(a_N - \log(1 + (\phi/r)\log(s/\omega_0))/\phi\big) \\ &\times \frac{\{\prod_{k=1}^{N-1}(1 - p(s_k, a_k))\}p(s, a_N)}{sr(1 + (\phi/r)\exp(\phi a_N)\log(s/\omega_0))} f_r(r)\,\mathrm{d}r \Bigg]. \end{aligned} \qquad (6)$$

The lower limit of integration can be replaced by $r_0 = \phi\log(s/\omega_0)/(\exp(\phi a_N) - 1)$, as for $r < r_0$ a tumour originating at birth cannot reach size s by age a_N.

A.3. The pdf of detecting a tumour of size s_p at presentation time t

The tumour now originates at age a_0 and is undetected at N screenings with probability $\prod_{k=1}^{N}(1 - p(s_k, a_k))$, where now tumour size at the kth screening is

$$s_k = s_p\exp\{-(r/\phi)(\exp(-\phi a_k) - \exp(-\phi t))\}.$$

The woman presents at tumour size s_p with pdf $q(\log(s/\omega_0))$, which gives a pdf of $q(\log(s/\omega_0))\,\mathrm{d}\log(s)/\mathrm{d}(t - a_0)|_{a_0}$ of presenting at age t. We also require survival to age t. On integrating over all tumour growth-rates r, we obtain

$$\begin{aligned} g(s_p, t \mid \mathbf{a}) &= R(t)q(\log(s_p/\omega_0)) \\ &\times \int_0^{\infty} f_r(r)f_a(a_0)|\partial a_0/\partial s_p|_{t,r}\partial\log s_p/\partial t|_{a_0,r} \\ &\times \prod_{k=1}^{N}(1 - p(s_k, a_k))\,\mathrm{d}r. \end{aligned}$$

From equation (1),

$$\partial\log s_p/\partial t_{a_0,r} = r\exp(-\phi t),$$

and the product of Jacobians is

$$J_1 J_2 = \frac{r}{s(r + \phi \exp(\phi t) \log(s_p/\omega_0))}.$$

Finally,

$$g(s_p, t \mid \mathbf{a}) = R(t) q(\log(s_p/\omega_0)) \int_0^\infty f_r(r) \times f_a\big(t - \log(1 + (\phi/r)\log(s_p/\omega_0))/\phi\big) \times \frac{r \prod_{k=1}^N (1 - p(s_k, a_k))}{s(r + \phi \exp(\phi t) \log(s_p/\omega_0))} \mathrm{d}r. \quad (7)$$

In the absence of screening, in equation (7) $p(s, a) = 0$, giving

$$g(s_p, t \mid \mathbf{0}) = R(t) \int_0^\infty q(\log(s_p/\omega_0)) f_r(r) \times f_a\big(t - \log(1 + (\phi/r)\log(s_p/\omega_0))/\phi\big) \times \frac{r}{s(r + \phi \exp(\phi t) \log(s_p/\omega_0))} \mathrm{d}r, \quad (8)$$

which is needed in equation (5).

It is convenient to use the notation $f(s, r, \mathbf{a})$ and $g(s, t, r \mid \mathbf{a})$ to denote the integrands of f and g, so that

$$f(s, \mathbf{a}) = \int_0^\infty f_r(r) f(s, r, \mathbf{a}) \, \mathrm{d}r,$$

etc.

A.4. True and false interval cancers

Sometimes it has been judged by studying X-rays from original screenings that interval cancers are true, in that the tumour was too small to have been detected at the time of screening. For false interval cancers, there was with hindsight a visible tumour on the X-ray. This information can be included in the analysis, by replacing the likelihood function of an interval cancer by the likelihood of a true or false cancer.

For a true interval cancer, the tumour origination time $a_0 > a_N$, the age at last screening. From equation (1), the condition is

$$r \geqslant \frac{\phi \log(s_p/\omega_0)}{\exp(-\phi a_N) - \exp(-\phi t)}.$$

The likelihood for a true interval cancer is then

$$g_T(s_p, t \mid \mathbf{a}) = R(t) q(\log(s_p/\omega_0)) \times \int_{\frac{\phi \log(s_p/\omega_0)}{\exp(-\phi a_N) - \exp(-\phi t)}}^\infty f_r(r) \times f_a\big(t - \log(1 + (\phi/r)\log(s_p/\omega_0))/\phi\big) \times \frac{r \prod_{k=1}^N (1 - p(s_k, a_k))}{s(r + \phi \exp(\phi t) \log(s_p/\omega_0))} \mathrm{d}r. \quad (9)$$

For a false interval cancer, the expression for $g_F(s_p, t \mid \mathbf{a})$ has upper integration limit of $\phi \log(s_p/\omega_0)/(\exp(-\phi a_N) - \exp(-\phi t))$.

A.5. Adding survival after diagnosis to the likelihood function

The modification to the likelihood function to include death or survival after diagnosis is straightforward. The functions $f(s, r, \mathbf{a})$ and $g(s, t, r \mid \mathbf{a})$ are simply each multiplied by the pdf of death after a further time τ, e.g., $f(s, r, \mathbf{a}) \rightarrow f(s, r, \mathbf{a}) h(\tau) S(\tau)$, or if the woman is still living when follow-up ceases at time τ after diagnosis, $f(s, r, \mathbf{a}) \rightarrow f(s, r, \mathbf{a}) S(\tau)$. The hazard function h and survival function S are in general functions of tumour size at diagnosis s, age at diagnosis a, elapsed time from diagnosis τ, and rate of tumour growth r.

On adding the dependence of the hazard h_b on r (2 parameters), twice the log-likelihood (distributed as a chi-squared) increased by 16. Hence this dependence significantly improves the fit to data. Replacing the r factor by a scaling factor that could scale down the hazard of death for screen-detected tumours gave about the same improvement in fit. Finally, using both terms multiplied together, a further increase of 4.4 was obtained. This shows that much of the reduction in mortality from screen-detected tumours is due to their slower growth-rate, but that not all of the effect of detecting tumours at screening can be explained in this way. The model constructed without explicitly adding a factor of hazard scaling for screened cancers gives a 5-year survival rate of 73.4% for interval cancers and 87.3% for screen-detected cancers.

References

[1] N.E. Day and S.D. Walter, Simplified models of screening for chronic disease: estimation procedures from mass screening programmes, Biometrics 40 (1984) 1–14.

[2] S.W. Duffy, H.-H. Chen, L. Tabar and N.E. Day, Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase, Statistics in Medicine 14 (1995) 1531–1543.

[3] T. Hopkins and C. Phillips, *Numerical Methods in Practice Using the NAG Library* (Addison-Wesley, Wokingham, 1988).

[4] J.T.M. Jansen and J. Zoetelief, MBS: a model for risk benefit analysis of breast cancer screening, British Journal of Radiology 68 (1995) 141–149.

[5] A.E. Johnson and J. Shekdar, Interval cancers in the National Health Service breast screening programme, British Journal of Radiology 68 (1995) 862–869.

[6] G.J. Oortmarssen, J.D.F. Habbema, P.J. Maas, H.J. Koning, H.J.A. Collette, A.L.M. Verbeek, A.T. Geerts and K.T.N. Lubbe, A model for breast cancer screening, Cancer 66 (1990) 1601–1612.

[7] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vettering, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[8] M. Schwartz, An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease, Cancer 41 (1978) 1550–1564.

[9] M. Zelen and M. Feinleib, On the theory of screening for chronic diseases, Biometrika 56 (1969) 601–614.