ScHARR

Twenty-two short models about mammography: how model design choices drive predictions of the benefits of screening

Jason Madan[1]

[1] Health Economics and Decision Science, School of Health and Related Research, University of Sheffield

Corresponding author:
Jason Madan
Health Economics and Decision Science,
School of Health and Related Research
University of Sheffield
Regent Court
30 Regent Street
Sheffield
S1 4DA
UK
Tel: 0114 222 0827
Fax: 0114 222 4095
Email: j.madan@sheffield.ac.uk

This publication presents work in progress. This work was made possible by a Researcher Development Award from the NIHR Research Capacity Development programme. The author would also like to acknowledge advice and support from Dr. Jon Karnon (University of Adelaide), Ms. Lynda Wyld (University of Sheffield and the Hallamshire Hospital, Sheffield), and Dr. Matt Stevenson (University of Sheffield). Comments are welcome, and should be sent to the corresponding author.

Introduction

The work presented here forms part of a PhD on model-based evaluation of screening for breast cancer by mammography. There are shortcomings to trial-based approaches to such evaluations [1]. However, model-based evaluations will be influenced by the assumptions implicit in their structure. A systematic review of models revealed that the clinical implications of assumptions were not always made clear, and that there were aspects of the disease relevant to the impact of screening that had not been included in existing models [2]. This raised the question – to what extent can the validity of model-based mammography evaluations be improved by increasing their complexity and realism? To answer this question, a series of breast cancer models of increasing sophistication were developed and fitted to a common dataset. The results were combined with survival models appropriate to each model structure to give a range of estimates of the benefits of screening.

Alternative assumptions for breast cancer natural history models

To allow for the systematic construction of models of increasing complexity and realism, an analysis of the factors driving the impact of screening on breast cancer outcomes was carried out. This was based on key texts and discussions with a consultant surgical oncologist (Ms Lynda Wyld).

Model choice type 1 – Choice of prognostic factors.

There are a number of prognostic factors linked to breast cancer outcomes:

- Progression describes the extent to which cancer cells have spread within the body, and is a key characteristic of breast cancer natural history. This can be defined as non-invasive, localised invasive, regional metastasis (i.e. lymph node involvement) and distal metastasis.
- Tumour size is a factor that can be linked with outcome. Also, since the presence of a lump is frequently the symptom that leads to presentation with breast cancer, and larger tumours are more easily visible by mammography, tumour size is likely to be an important parameter in modelling the latency period of breast cancer and the potential of screening.
- The characteristics of a tumour will influence its behaviour, and may also be influence treatment choices. Tumour grade is often used as a factor in treatment decision-making, and is therefore reasonably commonly found in breast cancer datasets.
- Genetic markers are becoming increasingly important in characterising tumours and influencing treatment. The most established of these is ER/PR, which has become widely noted with the introduction of tamoxifen. With the introduction of herceptin, HER2 is also being recorded more often.

Model choice type 2 – Tumour growth model

Tumour growth cannot be observed commonly in humans in vivo. The choice of growth model will therefore not be directly settled by data. It will also determine how tumour latency is extrapolated from size. Exponential growth is a common assumption in breast cancer natural history models. Gompertz growth, which describes exponential growth with asymptotic limits, was assumed in some models. One refinement of the Gompertz model is the stepped Gompertz model [3]. This has not been represented in screening evaluation models, perhaps because of the number of parameters required. A solution is to use a function that approximates to a stepped Gompertz growth curve, and a possible example is linear growth ($S = \alpha t$). This can be generalised to log-linear growth ($S = \alpha t^{\beta}$). Figure 1 illustrates these alternatives.

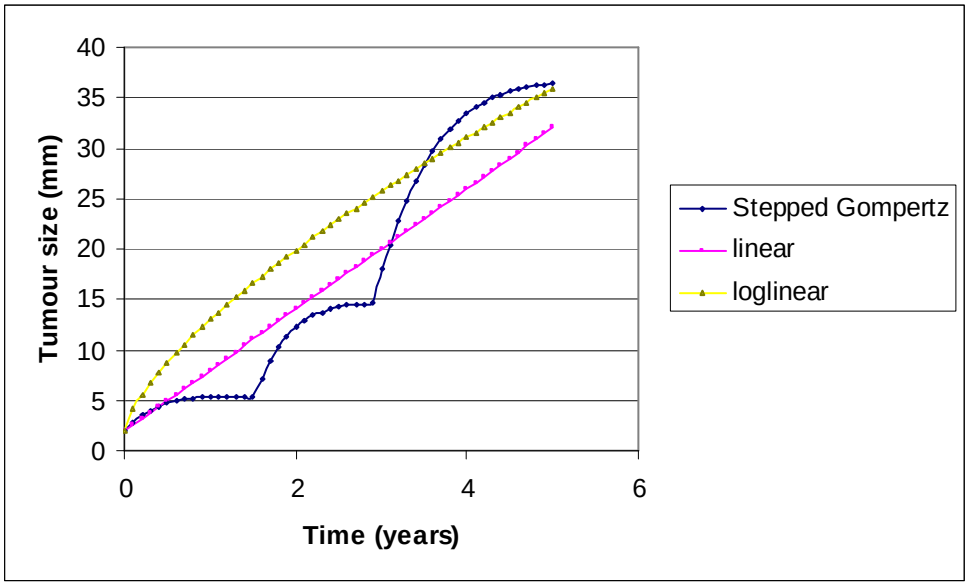


Figure 1 – An illustration of tumour growth models

Model choice type 3 – transition time distributions.

Usually, in fitting distributions that capture variation in transition times, data fit guides selection. However, there are a number of transitions in a screening model where dwell time is not directly observable. This makes the comparison of alternative distributions more challenging. The crudest state-transition structure is the Markov model, which uses exponential transition times. However, for most transitions in the model, the implied assumption of constant hazard lacks strong biological plausibility. Alternative viable distributions for transition times include the Weibull, gamma and lognormal distributions. These are illustrated in figure 2, below.

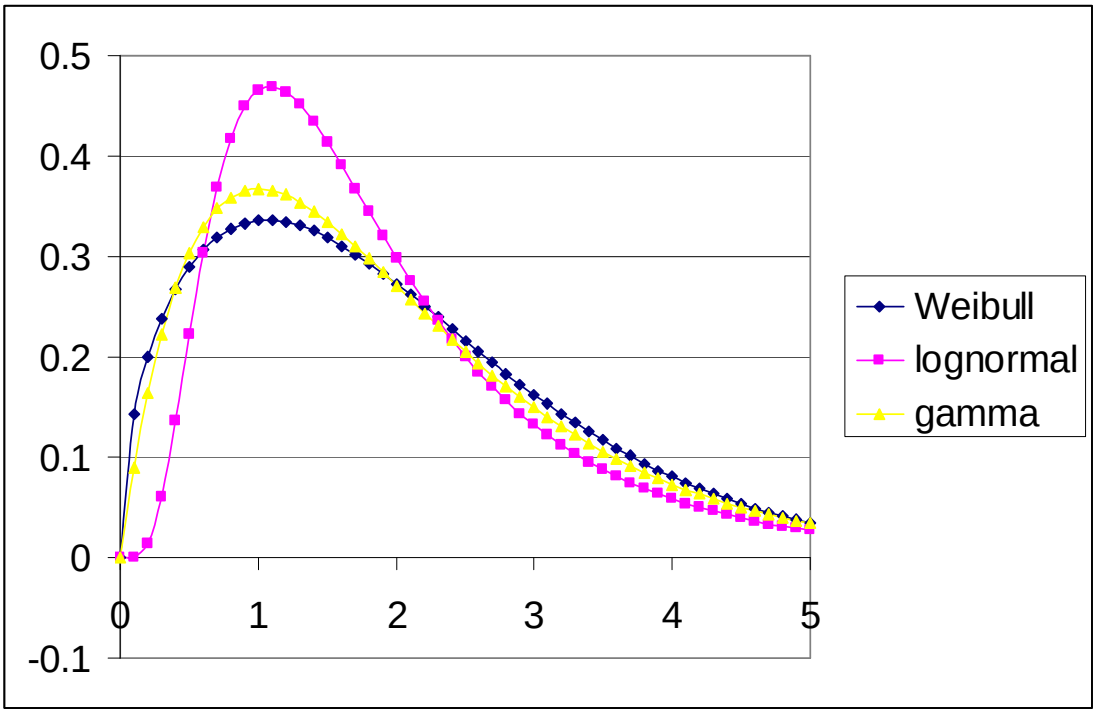


Figure 2: Alternative transition time distributions with mean and variance = 2.

Model choice type 3: Realism assumptions

This is a catch-all term for additions to a basic model structure based on insights into the underlying biology. For example, it has been known for some time that tumours are heterogeneous, and recent insights into the development of tumour cells and activities have enhanced this understanding[4]. One consequence of this is that progression through stages may well be related, so that a tumour that is adept at spreading to the lymphatic system will probably metastasise more quickly. It may well also be a faster-growing tumour. This can be represented in a model by inducing correlations between the relevant transition times. There is also evidence that tumours tend to be more aggressive in younger women, in that they are more likely to have metastases on presentation, and survival is poorer on average. Models can therefore be improved by making transition times age-dependent. A more sophisticated model would also allow for symptomatic presentation due to metastasis as well as tumour size.

Categorising alternative model structures

A large number of natural history model structures can be built from the choices identified above. One way to categorise these choices is that some influence the dataset to which the model can be fitted, and some do not. For example, introducing grade or ER status into a model requires data that records this information. However, alternative tumour growth models can be compared using the same dataset. Therefore, model structures can be grouped according to the datasets they can be fitted to. In order of increasing complexity, these groups are:

- Group one: stage progression models: These models only include the degree of progression as prognostic variables. The model structure choices are the type of density function for the transition times between stages, and the latency (the period during which the tumour is asymptomatic and detectable by mammography). The impact of inducing correlation between times can also be investigated.
- Group two: tumour growth models: These models include tumour size as the sole prognostic variable. The model structure choices are the type of tumour growth function and the density function for the latency.
- Group three: combined growth/progression models. Model structure choices for this group can be guided by results from groups one and two.
- Group four: grade models. These add histological grade, a marker of tumour aggression, as an additional prognostic variable. This allows correlations between transition times to be induced by making them grade-dependent.
- Group five: ER+ models. These add ER status, a marker for suitability for treatment with hormonal therapies and a prognostic factor.
- Group six: NPI models: These separate nodal involvement into two states, according to the number of nodes involved (1-3 or 4+).

Increasing the complexity of the natural history model brings a number of advantages. As complexity increases, the range of data that can be used for model fitting increases, as does the information that can be extracted. A broad range of prognostic information influences treatment choices. As the parameters of the model increase to simulate more and more of this range, the ability of the model to predict the impact of screening on treatment pathways improves. Perhaps most importantly, the structure of the natural history model will determine the realism with which survival can be modelled, so that more realistic models will produce more accurate estimates of the impact of screening on survival.

Datasets for model fitting

The aim is to gather a dataset with enough detail to allow the fitting and comparison of the alternative model structures described above. Two data sources stand out as being particularly useful for this purpose. They are the UK Breast Screening Programme (BSP) dataset, and the Breast Cancer Clinical Outcome Measures (BCCOM) project dataset. The former is collected by the NHS Cancer Screening centre. It gives the collated annual results for the UK BSP. For 2006-07, this involved results from screening 1.36 million women aged 50-70. 10036 cancers were detected in this group, and results are available segregated by age cohort, tumour size band and screening history.

The BCCOM presents results from an audit of non-screen detected cancers. It is less well established than the BSP data collection process, and only includes data on operable cancers (therefore, it excludes Stage III.ii and stage IV cancers). However, it provides information on a broader range of prognostic variables than the BSP (which only includes tumour size). The advantage of these datasets compared to trial data is that they are current, from a real-life UK context, and large. They also provide

sufficient detail for fitting the models described in this chapter. However, on their own, they are not quite sufficient for model fitting, and additional data is required.

An important variable that can be directly observed is the incidence rate. However, this is difficult to interpret in the context of mass screening programmes which artificially raise incidence. Data was obtained from the NHS in Scotland on incidence during the period 1980-88, which predates the BSP (English data was not readily available). This is shown in table A3.

Breast cancer models in the literature often draw heavily on data from the US National Cancer Institute Surveillance Epidemiology and End Results (SEER) datasets. This provides data on the proportion of cancers presenting with distal metastases, and the prevalence of ER positive cancers, as shown in table A4.

Finally, data from the Stockholm trial [5] was added to the dataset to allow for the synthesis of trial and routine data. This data was chosen as it related to first round (prevalence) screening specifically, and included information on progression and size. However, the Stockholm dataset is small relative to the other datasets, and contributes little to the likelihood. This makes it difficult to estimate progression transition times to a meaningful level of precision. Similar data from the UK screening programme would have a much larger sample size; however, it is not available at present. Therefore, to allow estimation of these key parameters, a revised model fitting exercise was carried out in which the Stockholm progression data was multiplied by a factor of 100. This was to an extent arbitrary, but was chosen to make the impact of the trial data comparable to the routine data. It is interesting to note that even a large trial of mass screening does not provide the information needed to fit realistic breast cancer natural history models. This is a consequence of the fact that a mass screening programme will have no impact on survival for the vast majority of participants.

Modelling the impact of survival

The natural history models predict the impact of screening on the prognostic characteristics of tumours on discovery. The choice of prognostic markers in the natural history model will drive the type of survival model that can be used to extend this prediction to estimates of lives saved through screening. In this paper, survival models are described for group two, group three and group six model types.

Group two: size-based survival.

Models from group two include tumour size as the sole prognostic factor. The size of a tumour at a given time will be a function of its growth rate and of the time since inception. A larger cancer will have a faster growth rate and/or been in existence for longer, and would therefore be expected to have a worse prognosis. This is the justification for the approach taken by Tan et al [6], who include in their model a fatal diameter parameter. In their model, the authors assume that a tumour becomes fatal with zero possibility of cure once it grows beyond this size. The fatal diameter is assumed to vary between individuals and is sampled from a Weibull distribution. The impact of improvements to treatment over time is included by changing the parameters of the Weibull distribution so that the mean fatal diameter increases

(shifting the distribution rightwards). The fatal diameter distribution for 1975 (in cm) has shape 0.95 and scale 4. The survival time from the point at which a tumour reaches fatal size is sampled from a lognormal distribution with the log of survival time (in years) having mean 2.4 and standard deviation 1.1. This information can be used with level two models to estimate the level of breast cancer mortality with or without any screening strategy.

Group three: Age/Stage/Size survival

Combining tumour growth and disease progression allows for the use of a more sophisticated survival model. Data is available from the US (SEER) giving 20-year mortality for combinations of age (30-49,50-69,70+), tumour size (<2cm,2-5cm,>5cm) and metastatic status (local,regional,distal) [7]. The data was of the form (D_{ijkl},N_{ijkl}), where:

i = year post-presentation (1-20)
j = age cohort (0 = 30-49, 1=50-69,2=70+)
k = size band (0=<2cm, 1 = 2-5cm, 2 = >5cm)
l = metastatic status (0 = local, 1 = regional, 2 = distal)
D_{ijkl} = the number of deaths recorded in year i amongst patients in age band j, size band k and metastatic status l.
N_{ijkl} = the number of patients alive at the start of year i in age band j, size band k and metastatic status l.

The data is binomial:

$D_{ijkl} \sim \text{binomial}(p_{ijkl}, N_{ijkl})$

where p_{ijkl} is the probability of dying in year i conditional on being alive at time i-1. It is therefore given by the equation

$$p_{ijkl} = S(i-1)-S(i) / S(i-1)$$

where S(t) is the survival function, i.e. the probability of being alive at time t. Rearranging gives

$$S(i)/S(i-1) = 1 - p_{ijkl}$$

$$\Lambda_{jkl}(i-1) - \Lambda_{jkl}(i) = \log(1 - p_{ijkl})$$

where $\Lambda_{jkl}(t) = -\log(S_{jkl}(t))$ is the cumulative (integrated) hazard function.

A number of survival models were fitted to this data. These models were fitted using Bayesian methods in the software package WinBugs.

Model one: Weibull survival

This model assumes a log-linear hazard. The cumulative hazard is of the form:

$\Lambda(t) = \exp(\mu)t^{\alpha}$

A proportional hazards approach was adopted, in which μ is a function of the prognostic variables:

$\mu_{jkl} = \mu_{base} + j*\beta_{age} + k*\beta_{size} + l*\beta_{prog}$, $l = 0,1$

$\mu_{jkl} = \mu_{base} + (j+1)*\beta_{mets}$, $l = 2$

The fitted values for these parameters are given in table 6.2.

node	mean	sd
alpha	0.8346	6.52E-03
betaage	0.03728	1.39E-02
betamets	1.196	1.29E-02
betaprog	0.8719	1.98E-02
betasize	0.4293	0.01491
mubase	-3.869	0.03064

Table 1 – Weibull survival model

Model three – log-logistic survival.

The log-logistic survival function has the cumulative hazard function

$\Lambda(t) = \log(1 + (\exp(\mu)t)^{\alpha})$

A proportional hazards model was fitted to the data:

$\mu_{jkl} = \mu_{base} + j*\beta_{age} + k*\beta_{size} + l*\beta_{prog}$, $l = 0,1$

$\mu_{jkl} = \mu_{base} + (j+1)*\beta_{mets}$, $l = 2$

The fitted values for these parameters are given in table 6.3

node	mean	sd
alpha	1.078	0.008542
betaage	0.02841	0.01669
betamets	1.674	0.02056
betaprog	1.058	0.02306
betasize	0.5449	0.0172
mubase	-4.161	0.02773

Table 2 – Log-Logistic survival model

Model 3 – Mixed survival model

It was observed from the data that distal cancers had a different survival profile to local and regional cancers. The former had a monotonically decreasing hazard, for which a Weibull distribution might be appropriate, whereas the latter had a unimodal hazard as would be given by a log-logistic distribution. Accordingly a mixed model was fitted in line with this observation.

node	mean	sd
alpha	1.105	0.009323
betaage	0.09749	0.01635
betamets	0.09849	0.03229
betaprog	1.105	0.02329
betasize	0.5978	0.01734
gamma	0.6313	0.01341
mubase	-4.314	0.02875
mumetsbase	-0.843	0.07601

Table 3 – Mixed Weibull / log-logistic survival model

Model fit was assessed by calculating the root mean squared error for the number of deaths predicted by each model:

RMSE (Weibull) = 28.84
RMSE (Log-logistic) = 22.04
RMSE (mixed) = 16.56

This suggests that the mixed model provides the best fit to the data.

Group six: NPI-based survival models

The additional degree of prognostic information included in level six natural history models allows them to be combined with survival models based on the Nottingham Prognostic Index (NPI). Data has been published on the impact of NPI score on survival [8]. The NPI scoring system does not apply to distal cancers, so their survival needs to be estimated separately. For this analysis, the Weibull metastatic model given in table 6.4 was used. As before, the data has a binomial distribution, with N(t) the number of deaths in year t and r(t) the size of the population at the start of year t. The only prognostic factor in this case is the NPI group, which is a categorical variable in the range (1,6). Two models were fitted to the data – a Weibull and a log logistic model.

	Weibull		log-logistic	
node	mean	sd	mean	Sd
alpha80	1.256	0.05567	1.612	0.07078
alpha90	1.261	0.06277	1.416	0.06923
betaprog80	0.6009	0.04066	0.508	0.03422
betaprog90	0.6883	0.04176	0.5674	0.04145
mubase80	-4.64	0.1815	-3.415	0.09888
mubase90	-6.002	0.1944	-4.562	0.1577

Table 4: NPI survival model parameter values.

Again, the root mean square error was calculated to assess the goodness-of-fit of both models:

RMSE (Weibull) = 6.03
RMSE (log-logistic) = 5.90

The log-logistic therefore was a marginally better fit to the data.

Natural History model – fitting technique

As mentioned before, models were grouped according to the type of data they could be fitted to. For each group, a subset of the full dataset described above was identified that the model could be fitted to. For example, the NHS Scotland incidence data was used in the fitting of all model groups. The BSP dataset, by contrast, was not used in fitting group 1 models, as they did not use tumour size as a progression variable. For some cross-tabulated data, tables were aggregated across variables not included in the model (e.g. the BCCOM dataset was aggregated across grade for the model types that did not include it). The aim was to identify values for model parameters that maximised the combined likelihood of the dataset used for fitting.

This was challenging, for two reasons. Firstly, most variables in a cancer screening model are not directly observable. The datasets, therefore, are usually functions of combinations of parameters, which increases the complexity of the likelihood function. Secondly, the data is aggregated into categories (age ranges, size bands) rather than available as precise values. The impact of these two factors is that the likelihood function involves a series of multidimensional integrals that cannot be solved analytically.

This likelihood function can be estimated using numerical approximation. However, there is an alternative. The data given above is all categorical, and therefore have multinomial distributions. The likelihood is a function of the probabilities in these distributions, and these probabilities are complex functions of multidimensional integrals that cannot be solved analytically. However, they can be estimated via Monte Carlo simulation of a cohort of women based on model parameter values. This can produce reasonably a accurate estimate of the likelihood for a sufficiently large cohort. The advantage of the Monte Carlo method is that it easily adaptable to models of increasing complexity, it produces estimates more rapidly than numerical integration, and it is easier to verify that the estimates are correct. For the purposes of fitting the models described in this chapter, the Monte Carlo method was used. All Monte Carlo simulation was carried out on the University of Sheffield grid computing facility, using code for the statistical programming language R.

Results

Group one models

Six group 1 models were fitted. Models 1A-1C used exponential, Weibull and lognormal transition time distributions respectively. Models 1D-1E added metastatic symptomatic presentation, presentation occurred at the minimum of standard or metastatic presentation. Model 1F added correlation between transition times. The results from fitting these models to the dataset given in the appendix are shown in table 5.

Model	No. Of Parameters	negative log likelihood	Distribution - latency	Mean Latency	Distribution - Time to regional metastasis	Mean time to regional metastasis	Distribution - Time to distant metastasis	Mean time to distant metastasis	Mean time from distant to symptomatic	Sensitivity
A	6	126171.2	Exponential	14.4	Exponential	13.5	Exponential	59.2	NA	69.0%
B	9	125356.3	Weibull	4.0	Weibull	4.5	Weibull	3.4	NA	74.4%
C	9	125279.6	Lognormal	4.7	Lognormal	7.3	Lognormal	3.7	NA	75.5%
D	10	125305.0	Weibull	4.2	Weibull	4.7	Weibull	3.4	0.49	78.5%
E	10	125257.8	Lognormal	4.3	Lognormal	6.6	Lognormal	3.4	0.48	76.9%
F	11	125255.9	Lognormal	4.2	Lognormal	6.6	Lognormal	3.3	0.53	75.9%
Full model	30	125017.0								

Table 5: Results from fitting Group 1 models.

Source	Type	Actual	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f
BCCOM	Local	12651.0	12878.5	12605.2	12634.6	12549.9	12586.8	12573.9
	Regional	11241.0	11013.5	11286.8	11257.4	11342.1	11305.2	11318.1
SEER	local	17409.0	18047.0	17538.4	17640.3	17502.9	17485.5	17466.7
	regional	15579.0	15433.5	15703.9	15717.5	15818.5	15705.1	15722.1
	distal	4232.0	3739.6	3977.7	3862.2	3898.6	4029.4	4031.2
BSP 1st screen	50-52	1110.0	2054.6	1117.4	1250.3	1181.1	1222.6	1197.0
	53-54	140.0	206.2	103.7	118.0	109.8	114.2	111.6
	55-59	125.0	230.3	106.1	121.4	113.4	117.6	113.3
	60-64	77.0	120.0	50.1	57.8	54.0	55.6	53.8
	65-69	68.0	99.9	38.0	43.9	41.3	41.8	40.7
BSP repeat	53-54	475.0	518.6	618.7	617.3	628.7	635.1	633.0
screen	55-59	1923.0	1645.6	1868.6	1858.8	1936.9	1937.7	1896.4
	60-64	2192.0	1660.7	1812.6	1844.5	1900.0	1891.5	1867.1
	65-69	1639.0	1273.9	1344.1	1348.6	1400.5	1374.4	1369.1
	70	120.0	87.0	90.3	92.1	94.5	92.9	92.6
ISD Scotland	25-29	10.0	3.2	3.8	0.4	2.7	1.7	1.8
cases	30-34	44.0	18.8	68.4	50.6	58.5	60.3	63.5
1980-88	35-39	96.0	45.2	133.6	116.4	124.4	125.9	127.8
Average per	40-44	167.0	76.6	178.3	158.0	167.7	170.2	172.8
year	45-49	245.0	114.6	220.2	205.0	208.8	212.4	215.7
and age	50-54	249.0	160.7	269.3	250.9	257.6	261.8	264.7
cohort	55-59	286.0	212.2	315.1	301.3	304.5	308.2	312.0
	60-64	301.0	259.1	349.5	333.0	341.9	346.0	343.2
	65-69	290.0	284.0	355.9	345.4	349.4	348.7	350.3
	70-74	295.0	307.9	359.6	350.9	354.2	353.9	356.6
	75-79	262.0	283.6	317.4	310.4	311.9	315.1	312.9
	80-84	187.0	209.9	222.0	217.0	219.3	217.3	219.4
Stockholm Trial	local	8525.0	7402.5	8440.2	8493.8	8460.7	8482.4	8452.9
(Up weighted	regional	2825.0	3586.1	2682.5	2707.9	2685.8	2717.5	2752.1
x100)	distal	25.0	386.4	252.3	173.3	228.5	175.1	170.0

Table 6: Actual and fitted values for group one models.

The negative log likelihood indicates how well each model fits the data; the lower the better, and the full model is the lowest possible. It is not intuitively obvious how closely the models fit the data from this measure. Table 6 gives actual and fitted models for the entire dataset. Figure 3 illustrates the probability density functions for latency and time to nodal involvement for each of the group one models.

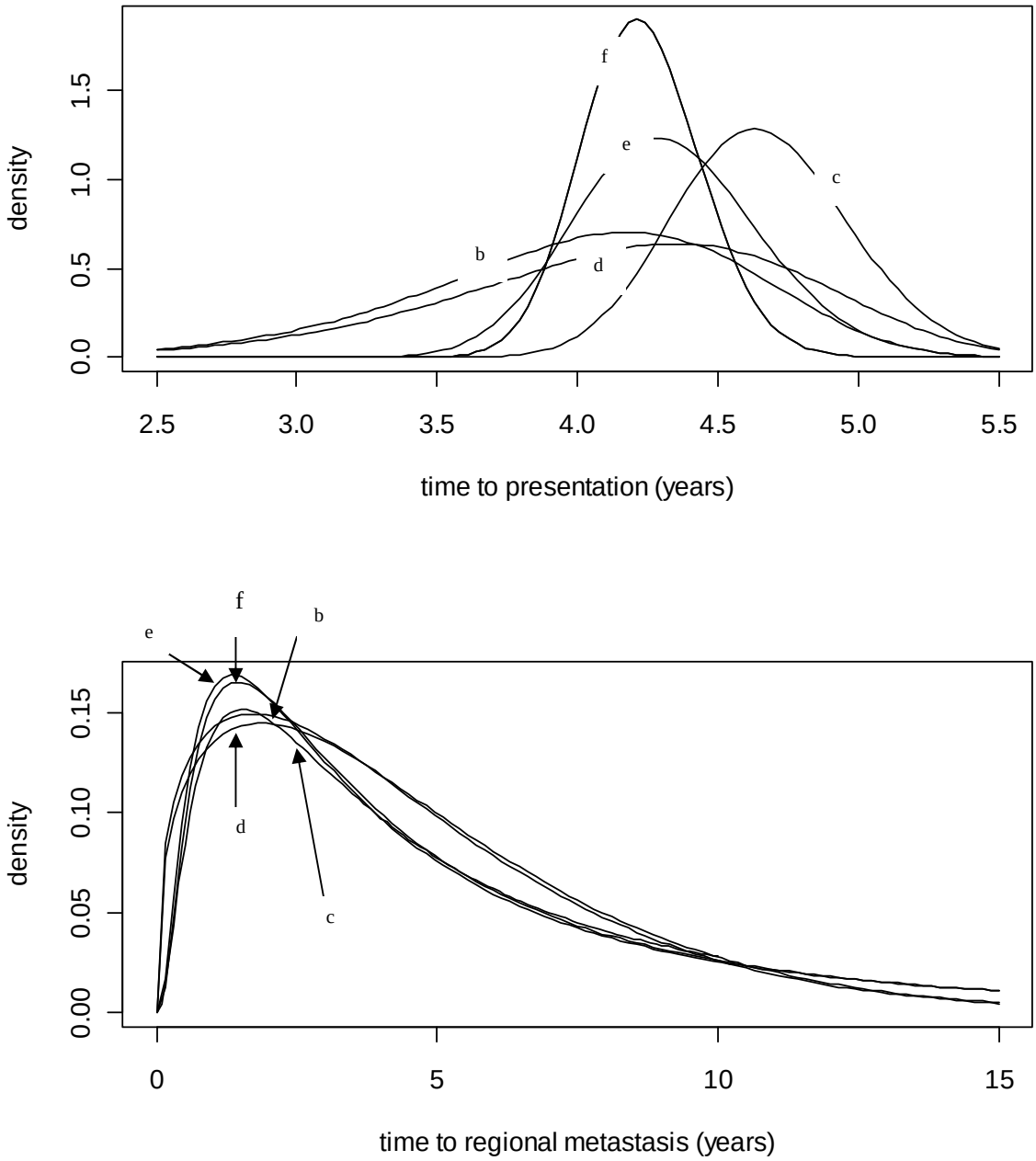


Figure 3: Transition time distributions for group one models

The results show that the standard Markovian assumption of exponential dwell times is a poor fit with the data, and the lognormal is a slightly better fit than the Weibull. Adding metastatic presentation and correlation between transition times does not appear to improve fit markedly for this dataset.

Group two models:

Seven tumour growth models were fitted. Model A assumed a constant tumour sensitivity (as per group one), and model B replaces this with a size-dependent sensitivity function Sens(S) = 1- exp(-α.S). Both assume exponentially distributed latency – models C and D replace this with Lognormal and Weibull distributions respectively. All four models assume exponential growth – models E-G fit alternative growth models. The results are given in table seven.

Model	No. Of Parameters	negative log likelihood	Growth model	Latency distribution	Mean Latency	Sensitivity (at 10mm)
A	6	141050.2	Exponential	Exponential	6.87	35.8%
B	7	138258.1	Exponential	Exponential	8.19	34.1%
C	7	137861.8	Exponential	Lognormal	13.16	20.0%
D	7	137988.7	Exponential	Weibull	8.17	29.1%
E	7	137475.6	Linear	Lognormal	2.97	83.0%
F	8	137299.8	Gompertz	Lognormal	3.39	87.3%
G	8	137308.4	Loglinear	Lognormal	3.17	90.4%
Full model	83	136511.2				

Table seven: Summary results from Type two models.

Again, exponential dwell times are a poor fit with the data, and the lognormal is a marginally better fit than the lognormal. The exponential growth model provides the weakest fit with the data, with the Gompertz and loglinear models almost equally good, and marginally better than the linear growth model. Figure four shows the mean growth profile over time for each growth model. The exponential growth models have longer latency times and lower sensitivities for smaller tumours. This is because exponential growth implies a high number of small asymptomatic tumours will be present in a population relative to large ones, and it is difficult to reconcile this with observations from screening programmes. Figure five illustrates the impact of model structure on predicted latency time distributions.

Using the fitted values, a large cohort of women was simulated for each model. For each cohort, mortality was first estimated in the absence of screening. Mortality was calculated in terms of lives and life years lost. The impact of two screening strategies on mortality was then calculated – triennial screening from ages 51 to 69 (i.e. the current UK strategy) and annual screening from ages 50 to 70, using the size-based survival model described above. Table 8 shows that the choice of model affects both the impact of screening and the relative benefit of annual vs. triennial screening.

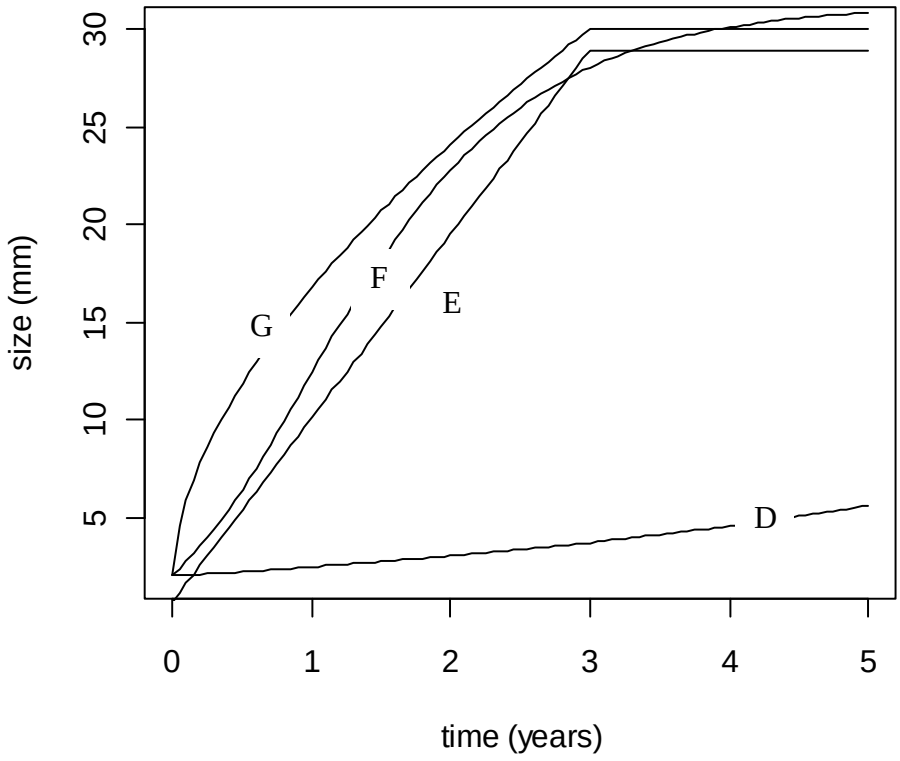


Figure 4: Mean growth paths for alternative group two tumour growth models.

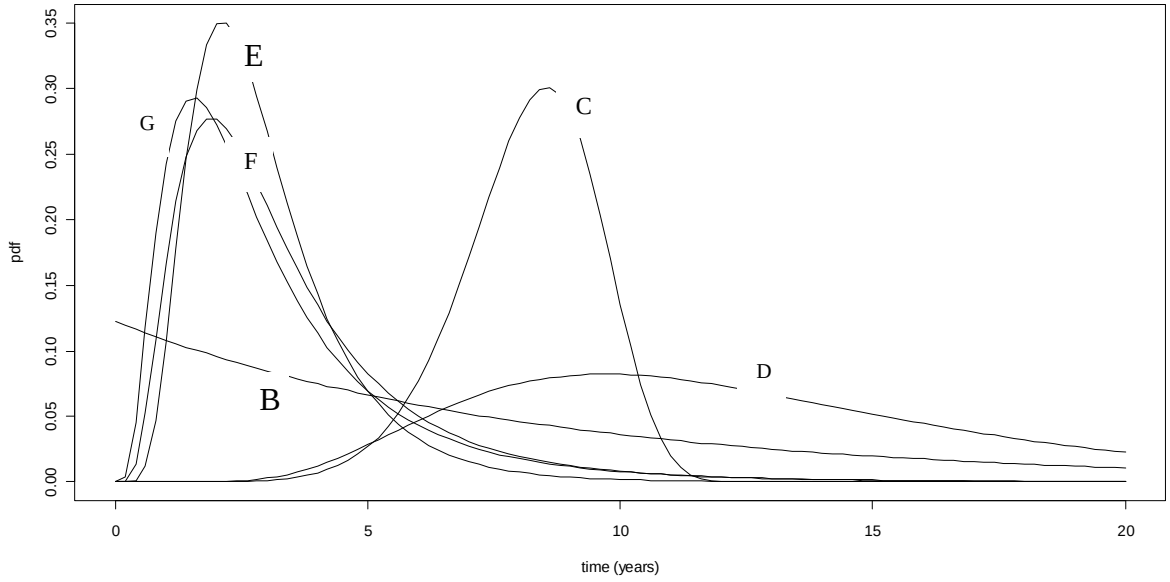


Figure 5: Latency period distributions for group two models B-G

	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2g
lives lost to breast cancer (no screening)	35,245	37,478	35,962	37,143	30,862	27,902	29,487
life years lost to breast cancer (no screening)	645,714	690,472	659,355	682,732	574,692	487,382	546,104
Lives saved (triennial screening)	5,101	844	784	841	898	866	798
Life years gained (triennial screening)	71,414	11,273	10,480	12,063	13,267	12,526	11,632
Lives saved (annual screening)	8,656	2,091	1,985	2,138	2,248	2,214	1,909
Life years gained (annual screening)	123,479	28,282	26,489	30,277	33,167	32,245	27,699

Table eight: Impact of model structure on breast cancer mortality (per million screened)

Results for models from groups three-six

Model	No of Parameters	Negative Log Likelihood	Tumour growth model	Sensitivity (at 10mm)	Mean latency	Mean time to node+	Mean time to metastasis	Additional features
3A	12	208411.7	Log-linear	94.7%	3.9	4.3	3.5	NA
3B	12	208342.0	Gompertz	90.3%	3.0	3.0	3.0	NA
3C	15	208091.5	Log-linear	98.5%	3.4	4.8	2.7	metastatic symptom and correlation between transition times
3D	15	208091.1	Gompertz	98.1%	3.1	5.2	3.1	metastatic symptom and correlation between transition times
3full	101	204250.0						
4A	18	233434.6	Log-linear	98.3%	5.3	6.4	8.8	Adds grade - transition times grade-dependent
4B	18	233397.4	Gompertz	98.0%	3.8	4.7	6.9	Adds grade - transition times grade-dependent
4full	116	227271.0						NB: Mean times for grade II tumour
5A	23	368113.8	Log-linear	87.5%	4.5	5.8	7.3	Adds ER status and latent tumour aggression parameter correlated with age, grade, ER and transition times
5B	23	367935.4	Gompertz	87.2%	4.1	5.3	5.3	Adds ER status and latent tumour aggression parameter correlated with age, grade, ER and transition times
5full	125	359798.0						NB: Mean times for woman aged 60
6A	25	383801.9	Log-linear	85.8%	3.2	5.1	4.2	Separates 1-3 nodes and 4+ nodes as separate states
6B	25	383265.8	Gompertz	82.1%	3.1	5.0	5.6	Separates 1-3 nodes and 4+ nodes as separate states
6full	130	372933.0						NB: Mean times for woman aged 60

Table nine: Results for model types 3-6

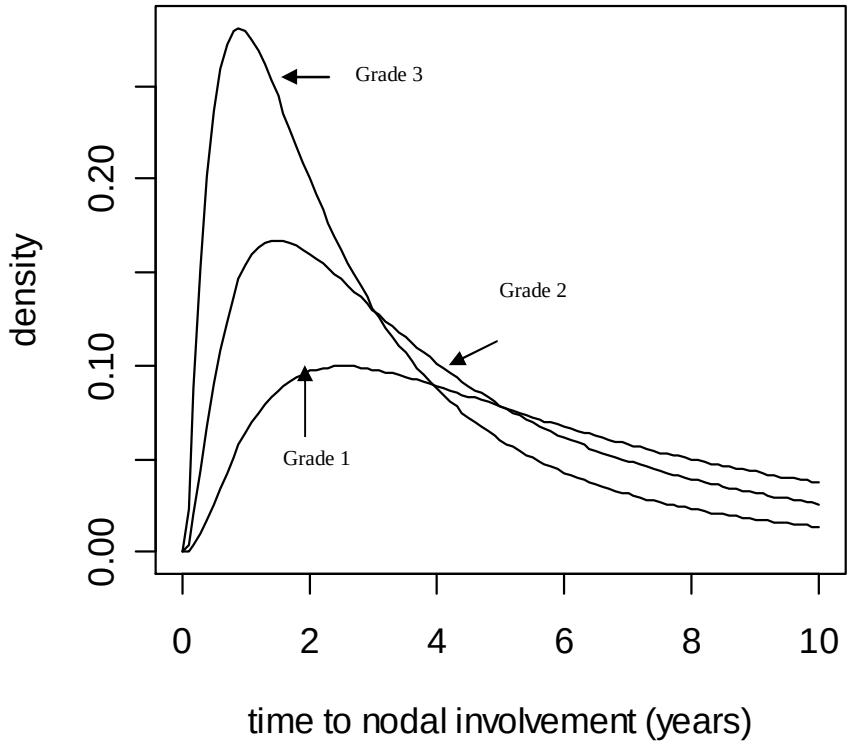
Based on insights from fitting group one and group two models, a number of choices were made for the more complex models. These were to model transition times using lognormal distributions, and to use the size-dependent tumour sensitivity function. Also, only log-linear and Gompertz growth functions were evaluated. Table nine shows the results for the more complex models.

	Model	3A	3B	3C	3D
Weibull Survival Model	Lives lost to breast cancer	52,337	65,419	55,038	54,081
	Life years lost to breast cancer	909,960	1,202,885	982,309	968,720
	Lives saved [triennial screening]	3,340	3,027	2,914	2,354
	Years saved [triennial screening]	63,345	59,122	58,291	48,689
	Lives saved [annual screening]	4,975	5,741	4,865	4,327
	Years saved [annual screening]	94,831	111,145	93,882	86,183
Log-logistic Survival Model	Lives lost to breast cancer	35,493	45,512	38,007	37,106
	Life years lost to breast cancer	598,239	807,872	656,876	642,255
	Lives saved [triennial screening]	3,289	3,299	2,931	2,551
	Years saved [triennial screening]	49,070	53,015	47,656	40,328
	Lives saved [annual screening]	4,887	6,270	5,092	4,608
	Years saved [annual screening]	71,006	101,848	80,194	71,517
Mixed Survival Model	Lives lost to breast cancer	37,536	47,967	40,145	39,222
	Life years lost to breast cancer	637,142	860,000	702,983	688,158
	Lives saved [triennial screening]	3,543	3,618	3,234	2,699
	Years saved [triennial screening]	56,847	58,764	56,940	49,069
	Lives saved [annual screening]	5,397	6,824	5,493	5,099
	Years saved [annual screening]	84,190	116,051	90,976	86,032

Table ten: Impact of natural history and survival model structure on benefit of screening

Group three models allowed for the fitting the age/size/progression models described previously. Once more, the impact of annual and triennial screening was compared; the results are shown in table ten. This shows that model structure for both natural history and survival affects predictions. The expectation was that including correlations between transition times would reduce the impact of screening, since screening is more likely to detect slow-growing tumours which are less likely to be fatal once correlations are induced. For the same reason, it was expected that inducing correlations would increase the relative benefit of annual against triennial screening. This is largely borne out in the results.

Group four models introduce grade and make the distributions grade-dependent (shown in figure 6 for time to nodal involvement). Survival analysis for the group 6 models, based on the NPI survival model described above, is currently being prepared.



Conclusions and Further Research

The work presented here is in progress, and some of the results may change. Further extensions to the modelling are required. The next step will be to include carcinoma in situ in the model, as one of the main results of mass population screening has been an increase in this type of cancer. Also, the models could be extended to include HER2 expression, given the recent interest in this marker. Further work is also required to identify uncertainty around the fitted values presented. This is not

straightforward, given the complex nature of the likelihood function. One approach that has been used elsewhere is to assume approximate normality around the maximum likelihood estimators. A better approach, given that parameter uncertainty is required, and correlation is likely to exist between parameters, would be Bayesian estimation of joint posterior densities using MCMC methods.

With the caveat in mind that the results are not final, they do show the impact of model structure on estimates of the benefits of screening. In particular, the exercise suggests that commonly made assumptions such as exponential growth and exponential dwell times are not a good fit with the data. The approach of increasing complexity step-wise allows simpler models to inform more complex ones and aids identification of all alternative structures. Also, the approach illustrates the link between the choice of model structure and the ability to model treatment and survival.

References

1. Madan J. From evidence-based to decision-analytic medicine: A mammography case study. University of Sheffield HEDS Discussion Paper Series. Available at http://www.shef.ac.uk/content/1/c6/01/87/47/HEDS%20DP%200808.pdf
2. Madan J. The role of modelling in the evaluation of breast cancer screening. PhD Thesis, in preparation.
3. Speer,J.F., Petrosky,V.E., Retsky,M.W., Wardwell,R.H. (1984). A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Research *44*, 4124-4130.
4. Baum,M., Chaplain,M.A.J., Anderson,A.R.A., Douek,M., and Vaidya,J.S. (1999). Does breast cancer exist in a state of chaos? European Journal of Cancer *35*, 886-891
5. Frisell,J., Glas,U., Hellstrom,L., and Somell,A. (1986). Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. Breast Cancer Research & Treatment *8*, 45-54.
6. Tan,S.Y.G.L., van Oortmarssen,G.J., de Koning,H.J., Boer,R., and Habbema,J.D. (2006). The MISCAN-Fadia continuous tumor growth model for breast cancer. Journal of the National Cancer Institute *Monographs.*, 56-65.
7. Cronin,K.A., Mariotto,A.B., Clarke,L.D., and Feuer,E.J. (2006). Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. Journal of the National Cancer Institute *Monographs.*, 26-29.
8. Blamey,R.W., Ellis,I.O., Pinder,S.E., Lee,A.H.S., Macmillan,R.D., Morgan,D.A.L., Robertson,J.F.R., Mitchell,M.J., Ball,G.R., Haybittle,J.L., and Elston,C.W. (2007). Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. European Journal of Cancer *43*, 1548-1555.

Appendix: Data used in model fitting

Category one – first time screened

AGE	No. (SN) Screened	Cancers detected	Status known	CIS	0-10mm	10-15mm	15-20mm	20-50mm	>50mm
50-52	205557	1553	1501	428	252	303	200	294	24
53-54	17889	182	175	40	31	37	21	39	7
55-59	16709	162	155	35	27	31	20	38	4
60-64	7028	90	89	12	17	20	10	29	1
65-69	4869	87	81	17	17	18	14	15	0
70	581	11	11	3	2	1	1	3	1

Category two – repeat attendees

AGE	No. (SN) Screened	Cancers detected	Status Known	CIS	0-10mm	10-15mm	15-20mm	20-50mm	>50mm
53-54	140864	636	622	158	125	125	95	112	7
55-59	388808	2439	2375	503	470	541	345	467	49
60-64	338794	2727	2649	520	571	595	436	489	38
65-69	227803	2019	1972	371	409	493	303	379	17
70	14777	145	144	24	37	40	18	24	1

Table A1 Results from the 2006-2007 BSP

2004-2006

		Negative	Positive
Grade 1	<10mm	545	69
	10-19mm	1232	407
	20-49mm	590	453
	50mm	36	69
Grade 2	<10mm	599	153
	10-19mm	2731	1380
	20-49mm	2453	3003
	50mm	203	682
Grade 3	<10mm	234	146
	10-19mm	1552	1012
	20-49mm	2308	3130
	50mm	168	737

2004 NPI

	<50	50-64	65-79	80+
Unknown	584	701	666	384
PPG	715	670	640	182
MGP2	753	732	667	202
MGP1	710	740	762	201
GPG	439	528	575	189
EPG	179	233	211	39
	3380	3604	3521	1197

Table A2 BCCOM results.

	Population	Annual incidence
25-29	188127	10
30-34	175238	44
35-39	168676	96
40-44	153607	167
45-49	147541	245
50-54	148592	249
55-59	150026	286
60-64	145797	301
65-69	132927	290
70-74	121739	295
75-79	98007	262
80-84	63235	187

Table A3 Breast cancer incidence – Scotland 1980-1988

	local	regional	distal
less2cm	7248	3466	
2-4.9cm	8918	8902	
5+cm	1243	3211	
size not recorded			4232
	17409	15579	4232

	ER+PR+ (n = 98,463)			ER+PR- (n = 19,886)			ER-PR+ (n = 4,896)			ER-PR- (n = 31,930)		
Age	n	Column %	Row %	n	Column %	Row %	n	Column %	Row %	n	Column %	Row %
30–39	4,903	5	49.5	885	4.5	8.9	569	11.6	5.7	3,558	11.1	35.9
40–49	18,575	18.9	60.9	2,490	12.5	8.2	1,508	30.8	4.9	7,929	24.8	26
50–59	21,807	22.1	61.4	4,464	22.5	12.5	1,161	23.7	3.3	8,094	25.4	22.8
60–69	22,019	22.4	65.9	4,752	23.9	14.2	801	16.4	2.4	5,856	18.3	17.5
70–79	20,667	21	68.2	4,662	23.4	15.4	584	11.9	1.9	4,401	13.8	14.5
≥80	10,492	10.6	67.7	2,633	13.2	17	273	5.6	1.8	2,092	6.6	13.5
Tumor grade												
1	18,012	21.8	81.1	2,885	17.5	13	405	9.9	1.8	914	3.3	4.1
2	40,642	49.3	74.2	7,133	43.4	13	1,324	32.5	2.4	5,682	20.6	10.4
3	22,082	26.8	44.4	6,027	36.6	12.1	2,174	53.3	4.4	19,412	70.4	39.1
4	1,774	2.1	45	413	2.5	10.5	173	4.3	4.4	1,579	5.7	40.1
Unknown	15,953	-		3,428	-		820	-		4,343	-	

Table A4 SEER data

Age	Screened	2 - 10mm	10 - 20 mm	>20mm	CIS
40-45	6517	2	1	3	2
45-50	5530	5	3	3	3
50-55	5977	9	4	2	5
55-60	7030	18	17	6	1
60-65	7479	22	13	5	4

	Size:	<10mm	10 to 20mm	>20mm
Negative	32405			
CIS	15			
local		50	26	9
regional		6	12	10

Table A5: Results from the Stockholm mammography trial