Evaluation by Markov chain models of a non-randomised breast cancer screening programme in women aged under 50 years in Sweden

H H Chen, E Thurfjell, S W Duffy, L Tabar

Abstract

Study objective—To apply Markov chain models that have previously been used on data in randomised trials of breast cancer screening to data from an uncontrolled service screening programme; to compare results with those from a randomised trial.

Design—A service screening programme in Uppsala county, Sweden, inviting 25 660 women aged 39–49 to mammographic screening every 20 months, and the Swedish Two-County Trial inviting 19 844 women aged 40–49 to two yearly screening, compared with 15 604 women of the same age in an unscreened control group. Data were analysed using Markov chain models and quasi-likelihood estimation procedures.

Main results—The results with respect to parameters of disease progression were very similar between the two studies. Use of estimated progression rates to predict the effect on mortality ranged from a 10% to a 17% reduction in breast cancer mortality in the Two-County Study and predicted a 15% reduction in the Uppsala programme. These compare well with the observed mortality reduction of 13% in the Two-County Trial.

Conclusions—The screening in the Uppsala programme is likely to have a similar effect to that observed in the Two-County Trial. It is feasible to evaluate non-randomised service screening programmes using Markov chain models.

(*J Epidemiol Community Health* 1998;**52**:329–335)

Graduate Institute of Epidemiology, College of Public Health, Taiwan University, Taipei, Taiwan
H H Chen

Department of Diagnostic Radiology, Mammography section, University of Uppsala, Uppsala, Sweden
E Thurfjell

MRC Biostatistics Unit, Cambridge, United Kingdom
S W Duffy

Mammography Department, Central Hospital, Falun, Sweden
L Tabar

Correspondence to:
Stephen W Duffy, MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge, CB2 2SR.

An important issue in breast cancer screening is the question of screening women aged under 50 using mammography. This concern was raised because of the evidence from randomised trials indicating that breast cancer screening confers a 40% breast cancer mortality reduction in women over 50 years of age, whereas the reduction is considerably smaller in women aged under 50 years.[1–3] Some research suggests that a smaller benefit of screening women aged under 50 is because of a shorter sojourn time coupled with slightly lower sensitivity in this age group compared with women aged 50 or more.[4] Accordingly, Tabar *et al*[4] suggest a shorter screening interval, say one year, for women age 40–49 years. Data from the Swedish Two-County Trial suggest that a one year screening regimen would have cut by approximately 50% the incidence of advanced tumours (tumour size larger than 2 cm, with node involvement and with malignancy grade 3) compared with a three yearly regimen.[5]

The above findings were mainly based on evaluation of randomised trials. However, routine (non-randomised) breast screening services are now common and there are service screening programmes inviting women aged under 50. Two issues arise in connection with this: firstly, can the mathematical models used in the past to analyse data from randomised trials of screening be used to evaluate a service screening programme with no control group,[6–13] and do results from such a programme agree with those randomised trials?

Duffy *et al* have developed a three state Markov chain model to estimate sojourn time (MST).[6] When applied to data from the Swedish Two-County study, this indicated a sojourn time of three to four years for women aged 50 or more, and one to two years for women under 50.[4] Chen *et al* extended this to a multistate Markov chain model to assess the preclinical natural history of breast cancer, including tumour growth, malignancy grade changes, and regional lymph node spread, and further to evaluate the effects of different screening frequencies on outcomes.[5 7]

Results of this suggest more rapid progression with respect to size and node status, and a more pronounced tendency for malignancy grade to worsen in women aged under 50.[5 7 14 15] Chen *et al* also extend the model to the simultaneous estimation of MST, sensitivity and specificity and found that sensitivity and specificity were poorer in women aged 40–49 than in women aged 50 or older.[5 8]

The service screening programme in Uppsala, Sweden includes women aged under 50 and some details of the first two rounds of screening in this age group have been published.[15 16] In this paper, we apply various Markov chain models of breast tumour progression to data from the Uppsala programme and from the Swedish Two-County study, with the following aims: (1) to estimate tumour progression in the Uppsala data from the preclinical to the clinical phase, and in terms of tumour size, malignancy grade, and regional lymph node spread; (2) to assess the effect of routine breast cancer screening in Uppsala, and of different screening frequencies on reducing advanced tumours, and compare the results with

Table 1   Summary of data, times (months), and transition probabilities used for estimation of transition parameters of M-N Markov chain model in the Swedish Two-County (women aged 40–49) and Uppsala studies (aged 40–49) studies

Detection mode and node status	Uppsala Transition Number history (Time, state*)	Two-County Transition Number history (Time, state*)	Transition probabilities applied
1 First screen			
Negative cases	25 660 (540,0→0)	18 456 (540,0→0)	P_{00}
Prevalent cases			
without nodal involvement	31 (540,0→1)	31 (540,0→1)	P_{01}
with nodal involvement	6 (540,0→2)	6 (540,0→2)	P_{02}
Interval cancers			
(between 1st and 2nd screen)			
without nodal involvement	14 (TSL,0→3)	14 (TSL,0→3)	P_{03}
with nodal involvement	9 (TSL,0→4)	9 (TSL,0→4)	P_{0}
2 Second screen			
Negative cases	19 986 (24,0→0)	16 396 (24,0→0)	P_{00}
Prevalent cases			
without nodal involvement	36 (24,0→1)	35 (24,0→1)	P_{01}
with nodal involvement	5 (24,0→2)	10 (24,0→2)	P_{02}
Interval cancers			
(between 2nd and 3rd screen)			
without nodal involvement	18 (TSL,0→3)	12 (TSL,0→3)	P_{03}
with nodal involvement	11 (TSL,0→4)	10 (TSL,0→4)	P_{0}
3 Third screen			
Negative cases	15 673 (24,0→0)	14 437 (24,0→0)	P_{00}
Prevalent cases			
without nodal involvement	21 (24,0→1)	30 (24,0→1)	P_{01}
with nodal involvement	2 (24,0→2)	6 (24,0→2)	P_{02}
Interval cancers			
(between 3rd and 4th screen)			
without nodal involvement	7 (TSL,0→3)	15 (TSL,0→3)	P_{03}
with nodal involvement	12 (TSL,0→4)	10 (TSL,0→4)	P_{0}

*TSL=time since last negative screen. In the model the actual times to occurrence of interval cancers are used.

the expected effect had there been no screening; and (3) to examine whether the results from (1) and (2) are consistent with those from the Swedish Two-County Trial.

Methods
Breast cancer screening with mammography for women aged 39–74 years was started in Uppsala county, Sweden on 1 February 1988. Full details of the programme and study design are given elsewhere.[2][16] For 25 660 women aged 39–49 years, there have been four rounds of screening 1988 to 1995 with a 20 month average interval between screens. The total number of breast cancers among those invited to screening were 214, including 38 diagnosed at prevalent screen, 88 at subsequent screens, 73 interval cancers, and 15 cancers in women who did not attend screening.

Two view mammography was used at the first screen and single view thereafter except where the initial screen indicated mammographically dense tissue. Approximately one third of mammograms were double read, the remainder single read.

In the Two-County Study, 19 844 women (Active Study Population, ASP) were invited to single view mammography ever 24 months (on average) for eight years, and 15 604 women (Passive Study Population, PSP) were invited to a single screen, contemporaneously with the final screen of the ASP. In the age group 40–49, there are 256 cancers in the ASP and 162 in the PSP. Mammograms were single read.

For both studies, we used only data from the first three screens and the intervals immediately following them, as tumour data from subsequent screens and intervals were very sparse. Also, when data were stratified by node status, further losses (of those with node status unknown) occur. Table 1 shows the numbers of tumours at each screen and interval, tabulated by node status.

A Markov chain model is one in which subjects move at random and independently between states. The fundamental assumption of such a model is that if we know the state at given time, t, say, the probabilities of subsequent movements are independent of states before t. To illustrate these models, let us consider a five state model incorporating phase of tumour development (no detectable disease, preclinical but screen detectable disease, symptomatic clinical disease) and node status (negative or positive). We assume that spontaneous regression of disease is impossible, and that there is no observation of progression after clinical diagnosis. The latter is because once a tumour is diagnosed, it is excised and subsequent observation of its natural history is impossible.

The above example gives the instantaneous transition intensity matrix as:

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} -\lambda_{01} & \lambda_{01} & 0 & 0 & 0 \\ 0 & -(\lambda_{12}+\lambda_{13}) & \lambda_{12} & \lambda_{13} & 0 \\ 0 & 0 & -\lambda_{24} & 0 & \lambda_{24} \\ 0 & 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

States: No disease (0), Pre-clinical N(−) (1), N(+) (2), Clinical N(−) (3), N(+) (4)

(1)

In the above matrix, λ_{ij} is the instantaneous transition rate from i to j—that is, the probability of moving from state i to j in an infinitesimal point in time. Thus, for example, λ_{01} is the screen detectable preclinical incidence rate, which represents the transition from no detectable disease to screen detectable preclinical node negative cancer in an instantaneous period of time. This parameter is, in theory, determined by a constellation of risk factors such as age at menarche, parity, and family

Table 2   Cancers tabulated by tumour attributes (tumour size, node status, and histology) and detection mode between the Uppsala and Two-County studies (rounds 1–3) in women aged 40–49

Detection mode and tumour attributes	Uppsala				Two-County			
	First screen	2–3 screens	Interval cancers	Refuser	First screen	2–3 screens	Interval cancers	Refuser
Tumour size								
In situ	7 (18.4)	16 (18.2)	6 (8.2)	0 (0)	8 (20.5)	13 (15.8)	5 (6.8)	0 (0)
1–9 mm	2 (5.3)	16 (18.2)	5 (6.9)	0 (0)	5 (12.8)	23 (28.0)	4 (5.5)	1 (16.7)
10–14 mm	10 (26.3)	22 (25.0)	11 (15.1)	8 (53.3)	11 (28.2)	18 (21.9)	14 (19.2)	1 (16.7)
15–19 mm	9 (23.7)	20 (22.7)	12 (16.4)	0 (0)	6 (15.4)	9 (11.0)	16 (21.9)	2 (33.3)
20–29 mm	5 (13.2)	11 (12.5)	27 (37.0)	3 (20.0)	3 (7.7)	12 (14.6)	17 (23.3)	0 (10.0)
30–49 mm	4 (10.5)	2 (2.3)	10 (13.7)	3 (20.0)	3 (7.7)	5 (6.1)	11 (15.1)	0 (0)
50 mm+	1 (2.6)	1 (1.1)	2 (2.7)	1 (6.7)	3 (7.7)	2 (2.4)	6 (8.2)	2 (33.3)
Total	38 (100)	88 (100)	73 (100)	15 (100)	39 (100)	82 (100)	73 (100)	6 (100)
Node status								
In situ	7 (18.9)	16 (16.7)	6 (6.9)	0 (0)	8 (21.6)	13 (16.0)	5 (7.1)	0 (0)
Node negative	24 (64.9)	55 (65.5)	34 (47.2)	8 (53.3)	23 (62.2)	52 (64.2)	36 (51.4)	3 (50.0)
Node positive and metastases	6 (16.2)	15 (17.9)	33 (45.8)	7 (46.7)	6 (16.2)	16 (19.8)	29 (41.4)	3 (50.0)
Total	37 (100)	86 (100)	73 (100)	15 (100)	37 (100)	81 (100)	70 (100)	6 (100)
Histology								
In situ	7 (18.4)	16 (18.2)	6 (8.2)	0 (0)	8 (20.5)	13 (16.3)	5 (7.0)	0 (0)
Ductal grade 1	2 (5.3)	13 (14.8)	6 (6.9)	2 (13.3)	6 (15.4)	17 (21.3)	4 (5.6)	0 (0.0)
Ductal grade 2	11 (28.9)	21 (23.9)	16 (21.9)	5 (33.3)	6 (15.4)	11 (13.8)	18 (25.3)	2 (33.3)
Ductal grade 3	13 (34.2)	22 (25.0)	30 (41.1)	6 (40.0)	11 (28.2)	22 (27.5)	30 (42.2)	2 (33.3)
Ductalgrade (NK)	0 (0.0)	0 (0.0)	1 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	2 (2.8)	1 (16.7)
Lobular	5 (13.2)	15 (17.1)	10 (13.7)	2 (13.3)	4 (10.3)	5 (6.2)	8 (11.3)	0 (0.0)
Medullary	0 (0.0)	0 (0.0)	3 (4.1)	0 (0.0)	0 (0.0)	3 (3.8)	2 (2.8)	1 (16.7)
Others	0 (0.0)	1 (1.1)	1 (1.4)	0 (0)	4 (10.3)	9 (11.2)	2 (2.8)	0 (0.0)
Total	38 (100)	88 (100)	73 (100)	15 (100)	39 (100)	80 (100)	71 (100)	6 (100)

Percentages are shown in parentheses.

history but we do not have information on these factors.

Standard mathematical procedures are available to convert the instantaneous rates into probabilities of transition in longer, non-instantaneous periods of time.[17] For example, in time t, the probability of passing from no detectable disease to preclinical disease is

$$1-\exp(-\lambda_{01}t)$$

Table 3   Estimates of parameters for the Uppsala (ages 39–49) and Two-County (ages 40–49) studies, using a three state Markov chain model

Parameter	Uppsala	Two-County
Pre-clinical to clinical disease (95% CI)	0.66 (0.57, 0.75)	0.41 (0.32, 0.56)
Mean sojourn time (y)	1.52	2.46
Sensitivity (%)	58	83

Table 4   One year transition probabilities for the Uppsala (ages 39–49) and Two-County (ages 40–49) studies, using five state Markov chain models related to tumour size (<2 cm, ≥2 cm), regional lymph node spread, and malignancy grade of ductal carcinoma

Tumour size

Initial states	→Final states	Uppsala	Two-County
1 Preclinical <2 cm	→Preclinical <2 cm	0.4314	0.5138
	→Preclinical ≥2 cm	0.0988	0.1533
	→Clinical <2 cm	0.2689	0.1766
	→Clinical ≥2 cm	0.2010	0.1562
2 Preclinical ≥2 cm	→Preclinical ≥2 cm	0.0957	0.2429
	→Clinical ≥2 cm	0.9043	0.7571
Regional lymph node spread			
Initial states	→Final states	Uppsala	Two-County
1 Preclinical <N(−)	→Preclinical N(−)	0.4494	0.5433
	→Preclinical N(+)	0.0642	0.0983
	→Clinical N(−)	0.2504	0.1954
	→Clinical N(+)	0.2360	0.1630
2 Preclinical ≥N(+)	→Preclinical N(+)	0.0250	0.1215
	→Clinical N(+)	0.9750	0.8785
Malignancy grade of ductal carcinoma			
Initial states	→Final states	Uppsala	Two-County
1 Preclinical g1/2	→Preclinical g 1/2	0.3215	0.3783
	→Preclinical g 3	0.2869	0.3083
	→Clinical g 1/2	0.1425	0.0447
	→Clinical g 3	0.2491	0.2686
2 Preclinical g3	→Preclinical g3	0.3185	0.3077
	→Clinical g3	0.6815	0.6923

Formulas for other transitions are more complex. Let us denote the transition probability matrix for time t as:

$$
P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{bmatrix}
P_{00} & P_{01} & P_{02} & P_{03} & P_{04} \\
0 & P_{11} & P_{12} & P_{13} & P_{14} \\
0 & 0 & P_{22} & P_{23} & P_{24} \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

with states: No disease (0); Pre-clinical N(−) (1), N(+) (2); Clinical N(−) (3), N(+) (4).

(2)

It should be noted that as each row includes all possible transitions from the antecedent state the summation of transition probabilities in each row is equal to 1. Note also that P_{ij} is a convenient shorthand for $P_{ij}(t)$. Table 1 illustrates the transitions observed in the two studies for the node status model.

The transition rates were converted into the transition probabilities in (2). To estimate the rates, a quasi-likelihood approach was adopted, using the SAS NLIN procedure. The elaboration of the likelihood function and estimation procedure are fully described elsewhere.[5–8] The results enable us to calculate the mean sojourn time (MST) and evaluate the effects of different screening frequencies by calculating the numbers of cancers by size, node status, etc, expected under each regimen. Details of the calculations can be seen in Chen *et al*.[5–7] Further details are available from authors.

We estimated the parameters of the following models in particular: (1) three state model for no detectable disease, preclinical screen-detectable disease and clinical disease; (2) five state model for preclinical/clinical phase and node status (M-N model); (3) five state model for preclinical/clinical phase and size <2 cm, 2+ cm (M-S-2 model); and (4) a five state model

for preclinical/clinical phase and malignancy grade 1–2, 3 (m-G-2 model).

Results

The overall distributions of tumour size, node status and histology with malignancy grade by detection mode (prevalent screen, subsequent screens, interval cancers) are similar between the Two-County trial (ASP, first three screening rounds) and the Uppsala study (table 2). Note that numbers vary slightly because size is almost invariably known, but node status and histological type are sometimes unknown. This table suggests that effectiveness of breast cancer screening in the Uppsala study is compatible with that obtained in the Two-County Study. However, there are fewer small cancers, tumour size 1–9 mm and ductal grade 1 screen detected cases, notably at the first screen, in the Uppsala programme.

Table 3 shows the estimated transition rates and MST from the three state model. Table 4 shows the results with respect to the five state models for node status, tumour size, and malignancy grade respectively, expressed as probabilities of transition within one year. Table 3 gives the MST in the Uppsala study as approximately 1.5 years compared with around 2.5 years in the Two-County trial. Results in table 4 suggest that it is more likely for preclinical tumours with size less than 2 cm or node negative to move into the corresponding clinical states than to the more advanced stage (larger than 2 cm or node positive) within the preclinical phase. The phenomenon is clearly seen for both size and node status in the Uppsala study, and for node status in the Two-County study. This is partly because of the fact that if a tumour does move to the more advanced stage, it has a very high probability of progressing thereafter to the clinical phase. For example, for a tumour that is presently node positive and preclinical, the probability of progressing to the clinical phase within one year is estimated as 0.975 from the Uppsala data and 0.8785 from the Two-County data. By con-

trast, there is a higher probability of progression from preclinical grade 1/2 to preclinical grade 3 than from preclinical grade1/2 to clinical grade1/2.

Table 5 shows the expected numbers of cancers by node status and detection mode, and by size (< 2 cm, 2 cm) and detection mode, calculated from the estimates for each study, by screening regimen (one year, two year, or three yearly screening). In each case, the expected numbers refer to a study group regularly screened for six years and a control group screened at the end of the six years, both groups comprising 20 000 women.

Results in table 5 suggest that when the screening interval increases there is an increase in the proportion of tumours with node status positive or larger than 2 cm in diameter. Thus, annual screening cuts by 40–50% the number of advanced tumours that would arise under a three yearly screening regimen. This result holds using estimates from both studies. To contrast different screening regimens with no screening, we take expected tumours in a control group in table 5 as the reference group, and calculate the relative incidence of advanced tumours for one yearly, two yearly, and three yearly regimens. Results are shown in table 6. The predicted effect of screening on malignancy grade is smaller than that predicted for tumour size or node status, for both centres

Table 5 Predicted cancers based on estimated parameters derived from the M-N (with and without node involvement) and M-S-2 (tumour size <2 cm, ≥2 cm) models, by detection mode and screening regimen assuming 100% attendance rate

Detection mode and screening regimens	M-N model				M-S-2 model			
	Uppsala		Two-County		Uppsala		Two-County	
	N(−)	N(+)	N(−)	N(+)	<2 cm	≥2 cm	<2 cm	≥2 cm
First screen								
One year	28.86	3.42	40.05	6.63	27.63	5.22	36.62	10.98
Two year	28.86	3.42	40.05	6.63	27.63	5.22	36.62	10.98
Three year	28.86	3.42	40.05	6.63	27.63	5.22	36.62	10.98
Later screens								
One year	94.84	8.83	109.09	11.27	93.76	11.87	106.22	16.07
Two year	68.66	7.57	84.11	11.67	67.04	11.13	80.33	18.12
Three year	52.09	6.00	66.72	10.20	50.44	9.07	62.82	16.48
Interval cancers (six years)								
One year	19.53	14.58	15.68	9.70	21.23	11.79	14.27	8.94
Two year	31.41	30.09	26.35	23.55	33.85	26.59	23.06	23.33
Three year	38.94	40.67	33.77	34.94	41.68	37.38	30.00	36.08
Total								
One year	143.23	26.84	164.82	27.61	142.62	28.88	157.11	35.99
Two year	128.94	41.08	150.51	41.86	128.51	42.94	140.61	52.43
Three year	119.88	50.10	140.53	51.78	119.75	51.67	129.45	63.54
Control before screen	62.76	75.28	62.49	83.62	65.67	73.24	52.96	96.91
Control at first screen	28.46	3.37	39.46	6.54	27.24	5.15	36.09	10.82
Total of the expected control								
group	91.22	78.65	101.95	90.15	92.91	78.39	101.95	90.15

Table 6 Relative risks of incidence of tumours with regional lymph node spread, larger than 2 cm in diameter and ductal grade 3 for one yearly, two yearly, and three yearly screening regimens in a six year period compared with an unscreened control group, estimated from the Two-County (ages 40–49) and Uppsala (ages 39–49) studies

	Uppsala			Two-County		
	M-N model	M-S-2 model	M-G-2 model	M-N model	M-S-2 model	M-G-2 model
Control	1.00	1.00	1.00	1.00	1.00	1.00
One yearly	0.34	0.37	0.48	0.31	0.35	0.44
Two yearly	0.52	0.55	0.66	0.46	0.51	0.61
Three yearly	0.64	0.66	0.75	0.57	0.61	0.72

and all screening regimens. Similar estimates are derived from each study.

It should be noted that as results in table 6 are based on a constant preclinical incidence rate, 100% attendance rate and 100% sensitivity, the relative benefits in reducing the rates of advanced tumours for the three screening regimens compared with the expected control group are too optimistic. Indeed, relative risks of node positive tumours or tumours larger than 2 cm for a two year regimen are lower than the observed relative incidences, 0.90 and 0.85 respectively in the Two-County study. This is probably because of the rapidly increasing incidence by age in this age group, poor sensitivity, non-attendance for screening, and excess diagnosis in the group invited to screening. We therefore calculate relative risks using an increasing underlying incidence rate estimated from Swedish national rates, adjusting for sensitivity, assuming 10%, 20%, and 30% refusal rates, and correcting for excess diagnosis by weighting according to the incidence rate of the control group (Two-County) or the preclinical incidence rate, λ_1 from the Markov chain model (Uppsala). These are translated into relative risks for death using the survival rates by tumour status from the Two-County trial.[7]

Table 7 shows the results. The predicted mortality reductions from two studies are similar. Predicted relative mortality and relative rates of advanced tumours for 10% refusal rates and a two year screening interval are close to those actually observed in the Two-County Trial.[15]

Discussion

In this project, we have shown that results on tumour progression and the effect of screening on rates of advanced cancers can be reproduced in analysis of a routine service programme. Using the Markov chain approach, we have estimated a 1.5-fold to threefold reduction (depending on the regimen) from screening in incidence of tumours with regional lymph node spread or size larger than 2 cm in diameter. The smaller benefit in reducing the proportion of poorly differentiated (grade 3) tumours suggests that the average 20 and 24 month intervals in the two studies might be too long to stop dedifferentiation occurring within the preclinical period. It should be noted that the screening benefit predicted from the Uppsala study is slightly lower than from the Two-County trial. This may be because of a higher interval cancer incidence rate as a percentage of the expected unscreened incidence rate in the Uppsala study, 16% and 50% for one and two years since negative screen, compared with 38% and 62% in the Two-County trial.[18] It should be noted, however, that five year survival rates were approximately 90% in this age group in Uppsala, similar to the rates observed in older women.[19]

There are two notable limitations in this project. The first caveat is that estimation of parameters for five state Markov chain models without adjusting for the sensitivity will be biased, as the sensitivity of mammography in

Table 7 Relative risks (95% CIs) of incidence of tumours with regional lymph node spread, diameter larger than 2 cm, and breast cancer death for one yearly, two yearly, and three yearly screening regimens in a six year period compared with an unscreened control group, estimated from the Two-County (ages 40–49) and Uppsala (ages 39–49) studies, adjusting for sensitivity and attendance, and estimating an underlying incidence which increases with age

	Uppsala				Two-County			
	M-N model		M-S-2 model		M-N model		M-S-2 model	
	RR node positive	RR Death	RR ≥2 cm	RR Death	RR node positive	RR Death	RR ≥2 cm	RR Death
Control	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
One year								
10%	0.59	0.75 (0.59, 0.96)	0.62	0.76 (0.60, 0.97)	0.51	0.75 (0.58, 0.97)	0.56	0.72 (0.54, 0.95)
20%	0.64	0.77 (0.60, 0.98)	0.67	0.78 (0.61, 1.00)	0.56	0.76 (0.59, 0.99)	0.63	0.76 (0.57, 1.00)
30%	0.69	0.79 (0.62, 1.01)	0.72	0.80 (0.63, 1.03)	0.61	0.78 (0.60, 1.01)	0.71	0.80 (0.61, 1.05)
Two year								
10%	0.77	0.85 (0.67, 1.08)	0.79	0.87 (0.68, 1.10)	0.74	0.90 (0.70, 1.16)	0.73	0.83 (0.63, 1.09)
20%	0.80	0.86 (0.68, 1.10)	0.82	0.88 (0.69, 1.11)	0.79	0.90 (0.70, 1.16)	0.79	0.86 (0.65, 1.13)
30%	0.83	0.87 (0.68, 1.11)	0.85	0.89 (0.69, 1.13)	0.90	1.01 (0.77, 1.30)	0.85	0.89 (0.68, 1.17)
Three year								
10%	0.86	0.90 (0.71, 1.15)	0.88	0.92 (0.72, 1.17)	0.89	0.98 (0.76, 1.26)	0.84	0.90 (0.68, 1.18)
20%	0.88	0.91 (0.72, 1.16)	0.90	0.92 (0.72, 1.17)	0.90	0.99 (0.77, 1.27)	0.88	0.92 (0.70, 1.21)
30%	0.91	0.91 (0.72, 1.16)	0.92	0.93 (0.73, 1.18)	0.91	1.00 (0.79, 1.29)	0.93	0.94 (0.72, 1.23)

Table 8 Comparison of estimated transition parameters (95% CIs) for the M-N model with and without adjustment for sensitivity in the Two-County (ages 40–49) and Uppsala (ages 39–49) studies

Studies Parameters of transition hazard rates	Uppsala		Two-County	
	Unadjusted	Adjusted	Unadjusted	Adjusted
λ_{01}: no disease to preclinical node (−)	0.00116	0.00116	0.00122	0.00122
λ_{12}: pre-clinical node (−) to preclinical node (+)	0.436 (0.240, 0.632)	0.417 (0.181, 0.653)	0.349 (0.194, 0.504)	0.366 (0.199, 0.533)
λ_{13}: pre-clinical node (−) to clinical node (−)	0.363 (0.175, 0.552)	0.384 (0.151, 0.617)	0.261 (0.108, 0.414)	0.243 (0.07, 0.409)
λ_{24}: pre-clinical node (+) to clinical node (+)	3.681 (0.577, 6.785)	3.55 (0.098, 7.000)	2.11 (0.726, 3.490)	2.18 (0.723, 3.65)

detecting breast cancer cases might be highly associated with tumour attributes. For example, a small tumour is more likely to be missed at screen than a large tumour. There are problems of identifiability when false negative probabilities are integrated into the likelihood for five state Markov chain models. To correct for the bias we first estimate the proportion of tumour attributes in false negative cases using maximum likelihood, apply this proportion to predict missed cancers at each screen and re-estimate the Markov transition parameters. Details of the procedure are provided in the Appendix. Table 8 shows that results adjusting for sensitivity are not dissimilar to those not allowing for sensitivity.

Note also that predicted relative mortality for a two yearly screening regimen from both programmes, adjusting for sensitivity, and with a 10% refusal rate (table 7), is around that observed in the Two-county study (0.87).

Another shortcoming arises from the M-G-2 model. As it could be argued that the tumour population is probably a mixture of tumours with and without the potential to dedifferentiate, estimation of transition parameters should be based on a mixture of transition probabilities rather than a model in which all tumours are capable of dedifferentiation. Sparse data on grade, however, preclude us from applying a mixture model to the Uppsala data. This bias is likely to be small, as previous research suggests that 84% of tumours in women aged under 55 are susceptible to phenotypic drift.[14]

Most of the qualitative and quantitative conclusions—a rapid progression to clinical phase, considerable potential for dedifferentiation in the preclinical phase, a modest reduction in mortality, and a strong dependence of the effectiveness of screening on the interscreening interval—are very similar for both studies. This indicates reproducibility of the results and strongly suggests that it is feasible to evaluate non-randomised programmes using the Markov models.

In terms of the Uppsala programme itself, use of Markov chain models to evaluate routine breast cancer screening suggests that the efficacy from the routine breast cancer screening service in the Uppsala is comparable to that in the Two-County randomised trial. Sensitivity may be a problem, and the screening programme in this age group must make every effort to maintain high standards of quality and film reading. In terms of use of the quantitative techniques, the methods can cope with the absence of a control group, but as with other complex models, estimation is difficult if the number of tumours is small. Results from both the randomised and non-randomised data are very consistent, notably in terms of prediction of benefit. If one year screening interval is applied to women aged 40–49 years and there is 90% attendance there is a likely to be a 20–25% mortality reduction in women invited to screening compared with women not invited.

Appendix

Estimation of tumour progression adjusted for sensitivity

We illustrate the procedure of re-estimation in the M-N model.

(1) According to Bayes' theorem, sensitivity for cancers without node involvement (s_1) and for cancers with node involvement (s_2) are calculated using the following equation:

$$S_1 = \frac{p's}{p(1-s) + p's} \qquad (1)$$

$$S_2 = \frac{(1-p')s}{(1-p)(1-s) + s(1-p')} \qquad (2)$$

where p=proportion node negative in false negative cases;
p'=proportion node negative in screen detected cases; and
s=overall sensitivity.

It should be noted that the overall sensitivity (s) is derived from a three state Markov chain model allowing for false negative and positive cases. Estimation of the proportion of node negative cases p among the false negative cases is carried out by minimising the difference between the observed and expected numbers of cases by each detection mode. p' is obtained directly from the observed screen detected cases.

(2) Calculate the proportion of false negative cases, u(i) say, surfacing to clinical stage during year I of the interscreening interval using the estimated transition probability matrix.

(3) Calculate the missed cancers at screen by node status and time since last negative screen by multiplication of the missed cases by u(i).

(4) Remove the false negative cases from interval cancers using random sampling stratified by node status and time since last negative screen as indicated in (3).

(5) Re-estimate the transition parameters using the M-N model.

1 Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA* 1971;**215**:1777–85.
2 Tabar L, Fagerberg CJG, Gad A, *et al*. Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985;i:829–32.
3 Nystrom L, Rutqvist LE, Wall S, *et al*. Breast cancer screening with mammography: an overview of the Swedish randomised trials. *Lancet* 1993;**341**:973–8.
4 Tabar L, Fagerberg CJG, Chen HH, *et al*. Efficacy of breast cancer screening by age: New results from the Swedish two-county trial. *Cancer* 1995;**75**:2507–17.
5 Chen HH, Duffy SW, Tabar L, *et al*. Markov chain models for progression of breast cancer Part I:tumour attributes and the preclinical screen-detectable phase. *Journal of Epidemiology and Biostatistics* 1997;**2**:9–23.
6 Duffy SW, Chen HH, Tabar L, *et al*. 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. *Stat Med* 1995;**14**:1531–43.
7 Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer Part II: Prediction of outcomes for different screening regimes. *Journal of Epidemiology and Biostatistics* 1997;**2**:25–35.
8 Chen HH, Duffy SW, Tabar L. A Markov chain method to estimate the tumourprogression rate from preclinical to clinical phase, sensitivity, and positive predictive value for mammmography in breast cancer screening. *The Statistician* 1996;**45**:307–17.
9 Zelen M, Feinleib M. On the theory of screening for chronic disease. *Biometrics* 1969;**56**:601–14.
10 Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984;**40**:1–14.

11 Prorok PC. The theory of Periodic screening. II. Doubly bounded recurrence times and mean lead time and detection probability estimation. *Advances in Applied Probability* 1976;**8**:460–76.

12 Albert A, Gertman PM, Louis TA. Screening for the early detection of cancer. I. The temporal natural history of a progressive disease state. *Math Biosci* 1978;**40**:1–59.

13 Shwartz M. An analysis of the benefits of serial screening for breast cancer based on a mathematical model of the disease. *Cancer* 1978;**34**:77–86.

14 Tabar L, Fagerberg CJG, Chen HH, *et al*. Tumour progression, histology and grade of ductal carcinoma: prognosis and progression. *Int J Cancer* 1996;**66**:413–19.

15 Tabar L, Fagerberg CJG, Chen HH, *et al*. Screening for breast cancer in women aged under 50: mode of detection, incidence, fatality, and histology. *J Med Screen* 1995;**2**:94–8.

16 Thurfjell EL, Lindgren JAA. Population-based mammography screening in Swedish clinical practice: prevalence and incidence screening in Uppsala county. *Radiology* 1994;**193**:351–7.

17 Cox DR, Miller HD. *The theory of stochastic processes.* London: Methuen, 1965.

18 Tabar L, Faberbeg CJG, Day NE, *et al*. What is the optimal interval between mammographic screening examinations? An analysis is based on the latest results of the Swedish two-county breast cancer screening. *Br J Cancer* 1987;**55**:547–51.

19 Thurfjell EL, Lindgren JAA. Breast cancer survival rates with mammographic screening. Similar facourable rates for women younger and those older than 50 years. *Radiology* 1996;**201**:421–6.