Statistical Methods in Medical Research 2010; **19**: 507–527

Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening

Harald Weedon-Fekjær Department of Etiological Research, Cancer Registry of Norway, Norway, **Steinar Tretli** Cancer Registry of Norway, Norway and Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway and **Odd O Aalen** Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

As mammography screening aims to improve the prognosis through earlier detection/treatment, tumour progression and screening test sensitivity (STS) represent key parameters in the evaluation of screening programs. We will here study some methods for estimation of tumour progression and STS, and show how previously used methods can be combined and developed to utilise more of the data available in modern screening programs. Weedon-Fekjær *et al*. recently suggested a study design using interview data about time since previous screening to estimate tumour progression and STS in a stepwise Markov model. While useful, the approach does not utilise tumour size measurements, nor link tumour progression to tumour size. Hence, we will here propose formulas for estimating tumour progression and STS using a continuous tumour growth model. To estimate tumour progression and STS, tumour growth curves are followed from one screening to the next, and probabilities for all combinations of tumour sizes at repeated screening examinations calculated. Based on the probabilities for screening detection on subsequent screening examinations, maximum likelihood estimates are calculated. Applied to Norwegian data, the new approach gives similar results to previously published results based on interval data, confirming the earlier estimated large variation in tumour growth rates.

1 Introduction

Since mammography screenings aims to lower breast cancer mortality through earlier diagnosis and treatment, a key parameter of every mammography screening program is its ability to detect breast cancers early (Figure 1(a)), and the ability will depend on both screening test sensitivity[1] (STS; the ability of a given screening to find cancer in the detectable preclinical phase) and tumour progression rates. Hence, tumour progression estimates are important in both optimising and evaluating screening programs. In addition, estimates of tumour progression are useful in general breast cancer research. However, there are no large observation studies of breast cancer tumour growth following clinical breast cancer tumours in humans, as almost all detected breast cancer tumours are treated in populations with good cancer registries. There are some observational studies of patients that were initially overlooked at earlier mammograms,[2–5]

Address for correspondence: Harald Weedon-Fekjær, Cancer Registry of Norway, Institute of Population-based Cancer Research, P.O. Box 5313, Majorstuen, 0304 Oslo, Norway.
E-mail: harald.weedon-fekjaer@kreftregisteret.no

10.1177/0962280209359860

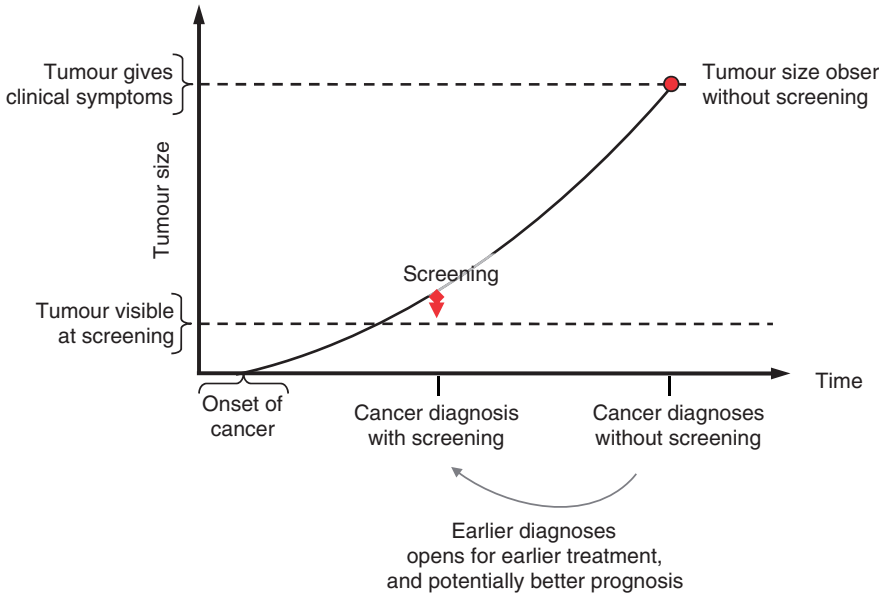

Figure 1 The idea of mammography screening: advance the diagnosis and the related treatment to a time before the cancer has spread to other organs.

or refused treatment,[4,5] but these studies are small, and likely influenced by length time bias, since slow-growing tumours spend relatively longer times in preclinical state visible on mammograms. Hence, STS and tumour progression have to be estimated indirectly by observed variations in cancer incidence due to screening programs. We will here review some models for tumour progression and STS estimation, from simple state wise Markov model to more complex continuous tumour growth models, and look at different ways of adopting the earlier approaches to study designs utilising more of the data available in modern screening programs like tumour size and time since last screening.

2 Current approaches to STS and tumour progression estimation

2.1 The basic stepwise models for estimation of tumour progression

Traditionally, tumour progression and STS have been estimated using state wise methods, mainly the MISCAN simulation model[6,7] or Markov models[8–11] Eventhough the estimation of parameters is quite different, the practical differences are relatively small as also the MISCAN model is based on the Markov property. The basic Markov model of cancer progression related to mammography screening is a three state model with women going from a state of 'no screening detectable cancer', through a state of 'pre-clinical cancer visible on mammograms', to a state of 'clinical cancer', as seen in Figure 2(a). Using this model, so called mean sojourn time (MST) and STS are estimated, where sojourn time is the pre-clinical time a cancer is visible on screening (Figure 3).

As cancer is assumed to be a gradually progressive disease, the definition and estimation of STS is not obvious. In practice, STS definitions vary, and interpretations of STS

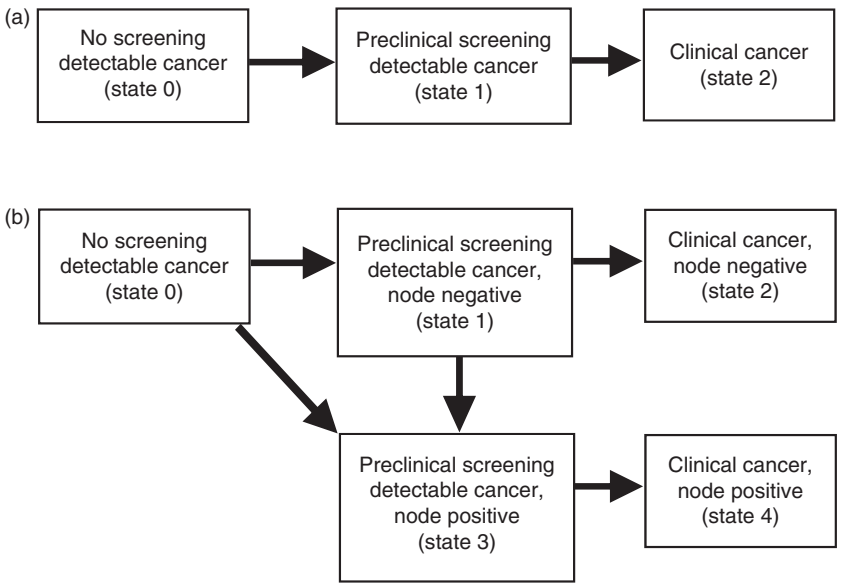

Figure 2 Examples of Markov models for breast cancer screening: (a) a simple Markov model for breast cancer screening and (b) a Markov model for breast cancer screening including axillary lymph node status.

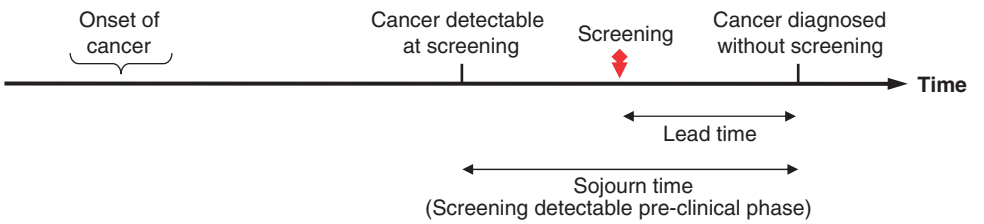

Figure 3 Illustration of terms related to screening evaluation, sojourn time and lead time.

must take into account the given definition. When monitoring screening programs, a common definition of STS is number of cancers found at screening divided by the total number of cancers found at screening and the subsequent year.[1,12] This definition is, however, questionable as a comparative measure between studies of different screening designs, as many cancers found on screening probably would have used several years to become clinically detected cancers without screening. Hence, STS in the Markov model is defined as the proportion of cancers in pre-clinical screening detectable phase that is detected at one screening round.

Usually, MST and STS have been estimated based on the observed number of cancers at screening and subsequent interval. Solving a special set of differential equations, so called Kolmogorov's equations, transition probabilities can be deduced. Based on the deduced transition probabilities, the probability of detecting cancers on screening, or in a given interval following screening, can be described as functions of STS, MST and the expected breast cancer incidence without screening.[8] Based on the functions for

probabilities of cancers at screening and the following interval, as given in Appendix 1, expected number of cases and estimates are deduced using either maximum likelihood estimation, non-linear weighted least square regression or Bayesian Gibbs sampling.[9,11]

In modern screening programs, other information is sometimes available, and women are usually followed for several subsequent screening rounds. In Weedon-Fekjær *et al.*[13] an alternative study design using interview data about time since previous screening combined with screening results was suggested, utilising the clear increasing trend in detected breast cancer cases on screening as a function of time since previous screening. Weedon-Fekjær *et al.*[13] deduced formulas for the probability of finding a breast cancer on screening as a function of time since previous screening (Appendix 1), and calculated MST and STS estimates.

2.2 Some problems with the `classical' Markov methods

In the Markov model, STS can be interpreted as 'the proportion of pre-clinical screening detectable cancers that is detected at one screening round'. While it is not dependent on screening design, it has one substantial weakness; there is no fixed definition of a cancer being in the state of 'pre-clinical cancers visible on mammograms'. In practice, a new improved screening method could actually move the state definition, extending the 'pre-clinical screening detectable phase', resulting in a lower STS even if the new screening procedure detects more cancers than the previous procedure.[14] Hence, parameters in the Markov model are not unambiguous 'biological' parameters that can be assumed to be equal for different screening tests, but to some degree local parameters more suitable for modelling the specific screening method. In addition to problems with the interpretation of STS, MST is correspondingly influenced by possible changes in state definition, increasing or decreasing the state of 'pre-clinical cancers visible on mammograms'. As both increases in the number of cancers found at screening, and decreases in the number of cancers seen in the interval following screening, which can be explained by either a higher MST or a higher STS, the two parameters are difficult to separate in simultaneous estimation. In practice, the estimation utilises minor contrasts in the data and estimates are highly correlated, as seen by their mutual confidence regions (Figure 4).

In addition to challenges regarding state definitions, the basic Markov model of breast cancer screening also has a limited way of dealing with individual variations in cancer progression. In the Markov model, transition times are assumed to follow an exponential distribution. Hence, cancer progression and individual variations in cancer progression only have a single variable, where mean and standard derivations of sojourn times are assumed to be equal. In practice, tumour progression is known to vary greatly between individuals,[2,15] and could probably be significantly larger than assumed in the Markov model. The stepwise models of breast cancer screening can be extended to non-exponentional distributions of transition times,[16] including increasing the possibility for individual variations as seen in random effect models,[17] but a more natural way are to use other models that utilise more of the available data like tumour size.

2.3 Advanced stepwise models of tumour progression

STS in the Markov model is modelled as a stepwise function, while the true STS are likely continuous function of tumour progression/size. To improve the modelling of

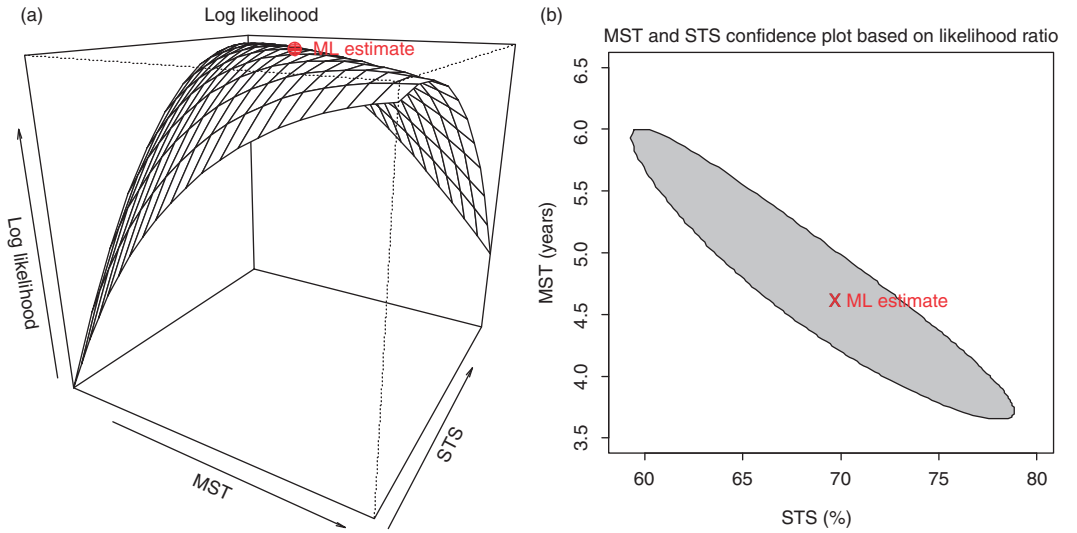

Figure 4 Illustrations of the strong correlation between MST and STS estimates derived from the 'classical' Markov model of mammography screening: (a) log likelihood of a Markov model, with a distinct ridge along a combination of increasing MST and decreasing STS (b) confidence region for MST and STS, with a distinct oval form. Data: NBCSP women 50–59 years of age, not corrected for increased HRT use.

the true disease process, several authors have extended the basic stepwise models with more states, as seen in Duffy *et al.* [9] where axillary lymph node status is added, as seen in Figure 2(b), for both the pre-clinical and clinical state. In these extended models, progression can follow different routes, and separate parameters for each transition are estimated. Fitting a simulation model to observed data by tuning parameters for best model fit, Oortmarssen *et al.*[7] used a model consisting of many states including three different size groups of invasive cancers. In these multi-stage models, STS and MST are defined for each state, hence reducing problems regarding interpretations of state definitions. Still, all transitions are assumed to follow an exponential, with a fixed ratio between tumour progression and variation in tumour progression.

2.4 Continuous tumour growth models

While the multi-state Markov models of tumour progression are likely to improve the representation of cancer progression, stepwise models do not fully take into account tumour size measurements found in several modern screening programs. As an alternative to the stepwise models, models based on continuous tumour growth have recently been suggested. Weedon-Fekjær *et al.*[15] used a continuous tumour growth model to estimate tumour growth and STS; and the US National Cancer Institute financed CISNET collaboration used continuous tumour growth models in their work to separate the effects of screening and improved treatment on breast cancer mortality rates.[18,19] With more data available, these models are likely to give more precise and biological relevant information regarding tumour progression than the earlier Markov models, including a separate variable for individual variations in tumour growth.

In the continuous tumour growth model of Weedon-Fekjær *et al.*[15] tumour growth was modelled based on a general logistic growth curve with a maximum size of 128mm in diameter, with a two parameter log-normal distributed variation in individual tumour growth rates. As for STS, it has a very different definition in the new model than in the Markov model, linking STS not to pre-clinical screening detectable phase, but directly to tumour size. Hence, STS estimates should not be directly compared between the two models. In practice, STS in the continuous tumour growth model was modelled using a two parameter logistic curve as a function of tumour size. Following the growth curves, using 'back calculation' from the expected number of future cancers, expected number of cases for a given parameter set was calculated and maximum likelihood estimates deduced. Applied to data from one screening examination in the Norwegian Breast Cancer Screening Program (NBCSP), and the following interval, tumour growth rate and STS as functions of tumour size were estimated. The estimates indicated a considerable variation in tumour growth rates, with 5% of tumours using less than 1.2 months to grow from 10 to 20mm in diameter, and another 5 % using more than 6.3 years. The mean time a tumour needed to grow from 10 to 20mm in diameter was estimated to 1.7 years, increasing with age. STS was estimated to increase sharply with tumour size, going from 26% at 5mm to 91% at 10mm. Compared to previously used Markov models for tumour progression, the applied model gave considerably higher model fit (85 % increased predictive power running cross-validations) and provided estimates directly linked to tumour size.

3 Norwegian screening data used in calculations

In 1995, the Norwegian government initiated an organised population based service screening program,[20,21] in which mammography results and interval cancer cases are carefully registered by the Cancer Registry of Norway. The NBCSP originally included four counties. Other counties were subsequently included, and by 2004, the screening program achieved nationwide coverage. All women between 50 and 69 years of age receive a written invitation biannually, and two-view mammograms are independently evaluated by two readers.

A high quality population based cancer registry and a unique personal identity number for each inhabitant in Norway enables close follow-up over time,[22] and the possibility to link data from several sources. Reporting of cancer cases to the cancer registry is mandatory, and information is obtained separately from clinicians, pathologists and death certificates.

The present study includes screening data from 1995 to 2002. A total of 78% of the invited women attended the screening program during this period, resulting in 364,731 screened women 50–69 years of age. Among these women, 336,533 answered a question regarding former screening experience, and 113,238 reported no previous (private or public) mammography experience before entering NBCSP. In the questionnaire, time since last mammography was given in the categories 0–1, 1–3, 3–5 and 5+ years. This format was not suitable for the estimation, so 0.67, 2, 4 and 6.5 years were chosen as appropriate representative points for each interval.

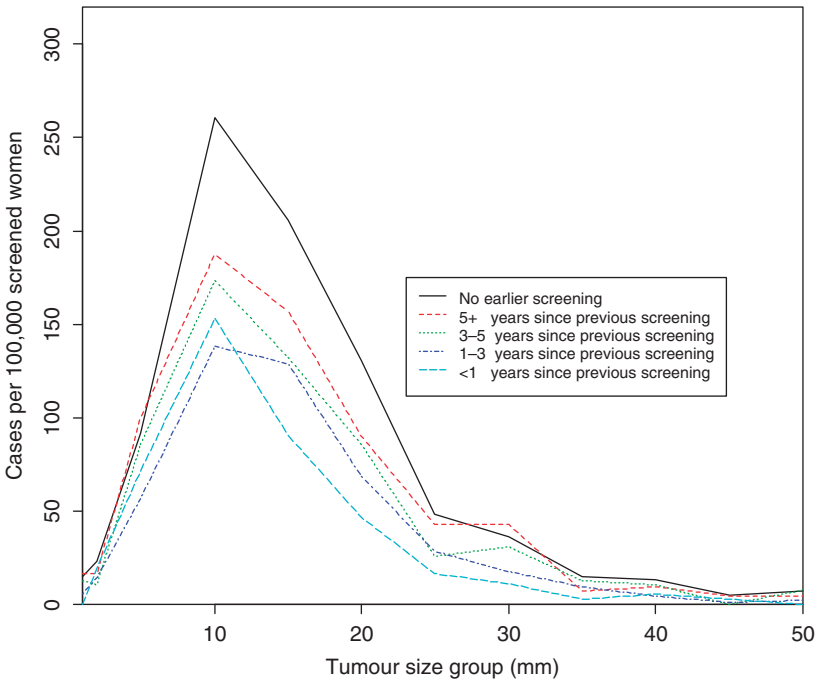

Figure 5 Distribution of NBCSP tumour sizes as a function of earlier screening, as used in the estimation of tumour progression and STS. Tumour diameter is grouped in 1, 2, 5, 10, 15, 20, 25 ... mm groups, and plotted up to 50 mm.

To make the results comparable to estimates provided in previous studies,[13–15] all cases of ductal carcinoma *in situ* (DCIS) were included. Several tumours detected at the same time in one woman were counted as one case, with size measurements given for the largest tumour. Only new primary breast cancers were included in this study.

In the NBCSP, tumour measurements are done on pathological lesions after surgery, and tumours are measured diagonally between the outer edges. All measurements were performed in a standardised manner according to specifications given in a national quality assurance manual.

Tumour size measurements of clinical breast cancers that emerge without screening are needed for the tumour growth model suggested in this article. Since women who do not attend screening represent a selected group, possibly with different alertness to early symptoms, tumour size measurements made before the start of the official screening program were used. The Cancer Registry of Norway did not receive reliable information on tumour size prior to the official screening program. However, at the Haukeland University Hospital, a good database for tumour measurements of clinical invasive breast cancers exists.[23] We were allowed to use these data, where 503 women aged 50–69 years were diagnosed with breast cancer between 1985 and 1994. Among these cases, 433 (86%) had registered tumour measurements in mm.

As the NBCSP offers screening to all women in the defined population, non-attending women are not considered a suitable control group due to possible selection effects.

Hence, background incidence was calculated from historical data combined with an estimated time trend. In practice, incidence data from 1990 to 1994 were used with time trend estimates from a special age-period-cohort model with additional screening parameters. This model has been described in detail by Møller *et al.*[24] Incidence rates vary among age groups and between counties, and the estimate was therefore weighted by the number of person years in each age group and county. As described in Weedon-Fekjær *et al.*[15] the assumed breast cancer incidence without screening is increased by 21% (when otherwise not noted) to account for the large increase in Norwegian HRT use during the study period.[25]

For an overview of data see Figure 5 and Table 1, and additional tables in Weedon-Fekjær *et al.*[15] and Weedon-Fekjær *et al.*[14] concerning interval data.

4 A continuous tumour growth model using time since previous screening

When estimating growth curves for an assumed continuous tumour growth model, Weedon-Fekjær *et al.*[15] included data from interval cancers combined with only the first initial screening round. We will here show how the method can be extended to cover study designs including repeated screening evaluations, and apply the new formulas to NBCSP data concerning one screening examination with varying time since previous screening.

As in Weedon-Fekjær *et al.*[15] tumour growth is expected to follow a logistic growth curve with parameter ¼ and a maximum tumour diameter of 128mm. This gives an almost exponentional growth for the smallest tumours, with decelerating growth as the tumours approach the maximum size of 128 mm in diameter. Further, individual tumour growth is assumed to be log normally distributed with mean α_1 and variance α_2, and STS is assumed to follow a two parameter logistic curve with one parameter defining how STS increases with increasing tumour size (β_1), and another linking a given STS to tumour size (β_2).

Working with screening data including time since previous screening, model parameters can be estimated by maximum likelihood. As the full likelihood includes several integrals, the actual maximum likelihood calculations are done discretely, grouping the data into sufficiently small time and tumour size intervals. Conditional on the assumed background incidence without screening and the clinical distribution of tumour sizes, the likelihood of a given dataset can be written as:

$$
\begin{aligned}
&L\left(\text{data} \mid \alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2\right) \\
&\quad \approx P(\text{observed number of cases at screening in different size intervals} \mid \text{no earlier screening and } \alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2) \\
&\qquad \cdot \prod_{\substack{\text{All observed times since} \\ \text{previous screening}}} P\left(\begin{array}{l} \text{observed number of cases at screening in different size} \\ \text{intervals} \mid x \text{ years since previous screening} \\ \text{and } \alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2 \end{array} \right),
\end{aligned}
$$

where each part is calculated by a multinomial distribution:

$$P(\text{observed number of cases at screening in different size groups} \mid \alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2) = \frac{sn!}{\prod_{i=1}^{sn} sc_i!} \cdot \prod_{i=1}^{sn} sp_i^{sc_i},$$

where i is an indicator for size group, sn the number of screened women, sc_i the number of screening cases in size group i, sp_i the probability of a woman having a tumour in size group i at screening given the parameter set $\{\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2\}$.

As seen in Weedon-Fekjær *et al.*[15] sp_i for the no previous screening group can be found by:

$$sp_i = \overbrace{S\left(\text{Cancer of size } i \mid \beta_1, \beta_2\right)}^{\text{STS for a tumour of size } i} \quad \overbrace{\sum_{\substack{\text{all time}\\ \text{intervals } f}} r \cdot gs_{f,\,i}}^{\substack{\text{Probability for a tumour of size } i,\\ \text{calculated using 'back calculation' from}\\ \text{expected future breast cancer rates}}},$$

where $S(\ldots)$ is STS defined by the logistic curve with parameters β_1 and β_2, r the expected breast cancer rate per time unit (month) without screening. To simplify calculations, the rate is assumed constant over time as in the earlier used Markov model,[8,11] probably giving a good approximation n the limited time span used in the estimation and $gs_{f,\,i}$, is the probability that a clinical cancer is in size group i, f months before clinical detection. Using our assumed tumour growth function $gs_{f,\,i}$ can be calculated using 'back calculation' of tumour sizes;

$$gs_{f,\,i} = \sum_{\substack{\text{all}\\ \text{sizegroups } g}} p_g \cdot P\left(\text{Tumor of size } g \text{ was of size } i,\ f \text{ months earlier} \mid \alpha_1,\ \alpha_2\right),$$

where p_g is the relative proportion of breast cancers of size g without screening. When extending the approach to a subsequent screening examination, it can be tempting to use the deduced probabilities of an initial screening, and add probabilities for screening detection x years earlier. This approach does not, however, take into account the large variation in tumour growth rates, and the increased probability of screening detection on subsequent screening examinations for slow growing tumours. Hence, the probabilities on subsequent screening $sp_i(x)$ given x years since previous screening must be calculated

looking at the whole time span and can be found by:

$$sp_i(x) = S\left(\text{Cancer of size } i \mid \beta_1,\ \beta_2\right) \cdot \sum_{\substack{\text{all time}\\ \text{intervals } f}} r \cdot gs_{f,\,\mathrm{i}} \cdot \overbrace{\sum_{\text{all sizegroups q}} gs_{f+x,\,\mathrm{q}} \cdot \left[1 - S\left(\text{Cancer of size } g \mid \beta_1,\ \beta_2\right)\right]}^{\substack{\text{Probability for tumour of size } g \text{ at previous screening,}\\ \text{combined with not being detected ar screening given size } g}}$$

where x is time (years) from previous screening.

As in Weedon-Fekjær *et al*.[15] the computations can be done by:

(1) Rearranging the growth formula expressing earlier tumour size as a function of present tumour size and tumour growth rate (κ_i).
(2) Calculating upper and lower limits for tumour growth (κ_i), constituting the requested probability.
(3) Calculating the probability for a tumour growth (κ_i) within the given limits using the log-normal distribution and assumed growth parameters $\{\alpha_1, \alpha_2\}$.

This is done for all relevant combinations of tumour size and future time intervals, projecting the tumour growth of a given future clinical cancer to the size groups at screening as illustrated in Figure 6(a). Finding the lower and upper growth rate bounds for a given size at screening, the probability for being in the given size group at screening can be calculated for a given set of growth parameters $\{\alpha_1, \alpha_2\}$. Extending the approach to subsequent screening as seen in Figure 6(b), growth rates are split by both size groups at present and former screening. In practice this is done by:

(1) Calculating tumour growth rates corresponding to borders of size groups at initial and subsequent screening examinations.
(2) Sorting the given growth rate limits by size.
(3) Calculating probability for detection on subsequent screening given initial screening in each groups, and summarise by size groups at subsequent screening.

4.1 Statistical calculations

All calculations, simulations and plots were done using the R statistical package.[26] Estimates were calculated using non-linear least square regression for the Markov model and maximum likelihood estimation for the continuous tumour growth model. To double check the implemented R functions for the continuous growth model, new data sets were simulated and results compared to the expected number of cases. For interval data, time intervals of one month were used. For the continuous growth model, tumour sizes were categorised to 1, 2, 5, 10, 15, … , 100 mm, as the background data revealed that many pathologists approximated tumour sizes to the nearest 5, 10, 15, … , 100 mm (data not shown).

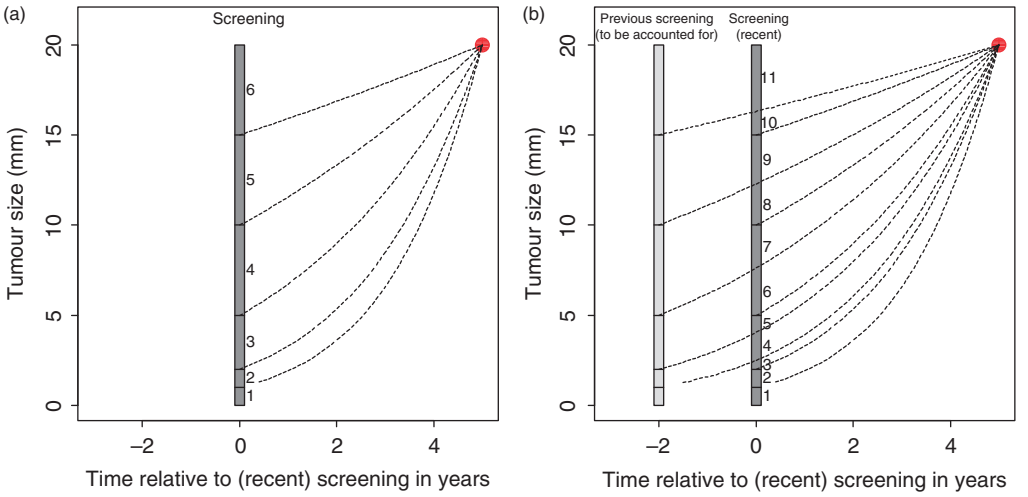

Figure 6 Illustration of the principle behind 'back-calculations' of tumour growth, used in the calculation of expected number of cases on repeated screening examinations. Based on potential future clinical cancers, growth curves are projected back to the observed screening examination, to find expected number of cases in each size group at screening; (a) projects are done for one initial screening (left) and (b) subsequent screening examinations (right), and probabilities for screening detection calculated for all combinations of growth curves at recent screening (marked with numbers).

The main estimates were given with (pointwise) confidence intervals showing their (random) uncertainty. Robust 95% confidence intervals were calculated by 100 bias corrected parametric bootstrap replications, resampling all the observed data except the assumed breast cancer incidence without screening.

5 Combining different approaches

In Weedon-Fekjær *et al.*[15] we suggested using screening data on time since previous screening as an alternative to interval data when estimating MST and STS under the Markov model. Similarly, we have here shown how data on time since previous screening can be used with a continuous tumour growth model. While data on time since previous screening is sometimes needed as there is no reliable interval data, these data can also be viewed as an extension of the basis for tumour progression and STS estimates. As the model assumptions are the same using the two study designs, interval and time since previous screening data, the two approaches can actually easily be combined into one simultaneous estimation using either the Markov model or continuously tumour growth model, using all the available data. We will here give combined estimates for both models, and discuss the results. In addition, a R script[26] for the combined Markov model is attached (Appendix 2). The program can perform estimation both with screening data only including time since previous screening, 'classical' screening and interval data, and all data combined. As seen in the script, weights in the non-linear least square regression are chosen in a special manner. In weighted mean square regression, weights are commonly calculated as a function of the expected standard deviation of each data point.

Table 1 Overview of key data used in the estimation: number of screened woman and cases, by reported time since previous screening

	Women (*n*)	Cases (*n*)	C/W*
No previous screening	113,238	972	1808
Time since previous screening			
0–1 years	36,507	154	895
1–3 years	84,449	405	1023
3–5 years	38,580	230	1244
> 5 years	42,091	289	1391
Expected incidence without screening	245 per 100,000 person year		

*Cases per 100,000 screened women.

With known standard deviations, this is normally a good choice. However, when estimating the standard deviation, this choice can be problematic if the estimated standard deviation is dependent on model parameters. For our Markov Model data, number of interval cases follows a Poisson distribution, with estimated standard derivation equalling the expected value (mean) of each data point. Working with weighted mean square regression in Markov Models, we have found that the standard weights sometimes could seriously bias our results towards parameters giving larger weights. To avoid such biases due to no independent weighting, weights are standardised so that the sum of weights are always equally one, as described in Weedon-Fekjær's PhD theses.[27] In addition, we have also chosen to equally weight screening and interval data as seen in Weedon-Fekjær's PhD thesis.[27]

When using data from many sources we may want to avoid assuming a given level of breast cancer without screening, and the program can also be used without assuming a given level of breast cancer without screening.[13]

6 Results

As we have seen in Table 2, Markov model estimation using all data, combining the earlier approaches using interval cancer data[14] and time since previous screening,[13] gives estimates that are a compromise between the two separate estimates. Overall, estimates using all data are somewhat closer to the estimate based on time since previous screening, than the one based on interval data. Rather surprisingly, bootstrap confidence intervals for the combined model are not always narrower than the earlier estimates based on either time since previous screening or interval data.

While there for the Markov model has been demonstrated considerable differences between estimates based on interval cancers and estimates based on time since previous screening,[13] the continuous tumour growth model gives more similar results using interval or time since previous screening data (Table 3 and Figure 7). Estimated tumour growth is somewhat slower with larger individual variation using time since previous screening data, than based on interval cases. As for STS, estimates are somewhat lower using time since previous screening data (Table 3 and Figure 8). The new tumour growth estimates using time since earlier screening, are moderately affected by the correction for

Table 2 MST and STS estimates using different parts of the NBCSP data

	MST estimate (years)		STS estimate (%)	
	50–59 years	60–69 years	50–59 years	60–69 years
Using interval data*	3.9 [3.2–4.2]**	5.0 [4.3–5.5]**	75 [70–82]**	85 [80–90]**
Using time since screening	5.6 [4.0–6.6]***	6.9 [5.5–7.8]***	55 [43–67]***	60 [49–71]***
Using all available data	5.4 [4.3–6.0]**	6.1 [5.1–6.8]**	55 [50–64]**	68 [64–77]**

* Excluding earlier screened women from screening data.
** 95% smoothed bias corrected bootstrap confidence interval*.
*** 95% bias corrected bootstrap confidence interval.
Remark: Results are taken from Weedon-Fekjær's[27] PhD thesis.

Table 3 Tumour growth and STS estimates using time since previous screening, compared with previously published estimates based on initial screening and interval cases.[15] All estimates based on Norwegian Breast Cancer Screening Program data for woman 50–69 years of age

	Estimate using time since previous screening	Estimate using all available data	Estimate using initial screening and interval cases[15]
Time used from 10 to 20 mm (years)			
Mean	2.1 [1.8, 2.3]	2.0 [1.8, 2.2]	1.7 [1.5, 1.8]
Standard deviation	2.2 [1.9, 2.4]	2.1 [2.0, 2.4]	2.2 [2.0, 2.4]
Volume doubling time at 15 mm (days)			
25th percentile	76 [65, 88]	68 [57, 79]	41 [31, 47]
50th percentile	152 [134, 170]	142 [125, 161]	99 [83, 107]
75th percentile	303 [262, 337]	296 [263, 331]	234 [202, 254]
Screening test sensitivity at			
5 mm	14 [12, 16]	14 [11, 16]	26.5 [22.5, 30.8]
10 mm	58 [51, 56]	60 [53, 66]	91.5 [87.6, 96.4]
15 mm	93 [89, 96]	93 [90, 96]	99.7 [99.5, 100.1]

increased HRT use (Table 4). Generally, estimates based on time since previous screening seems more stable, using the continuous tumour growth model than the Markov model. As for the Markov model, estimates using all data under the continuous tumour growth model become a compromise between the two separate approaches, and did not always result in narrower confidence intervals (Table 3).

While estimating separate STS's for present and previous screening gave wide confidence intervals and little information about the true values for the Markov model,[13] the estimates seem stable for the continuous tumour growth model (Table 5). From our estimates, STS in the first NBCSP screening round seems higher than STS of preceding mammography outside the official screening program.

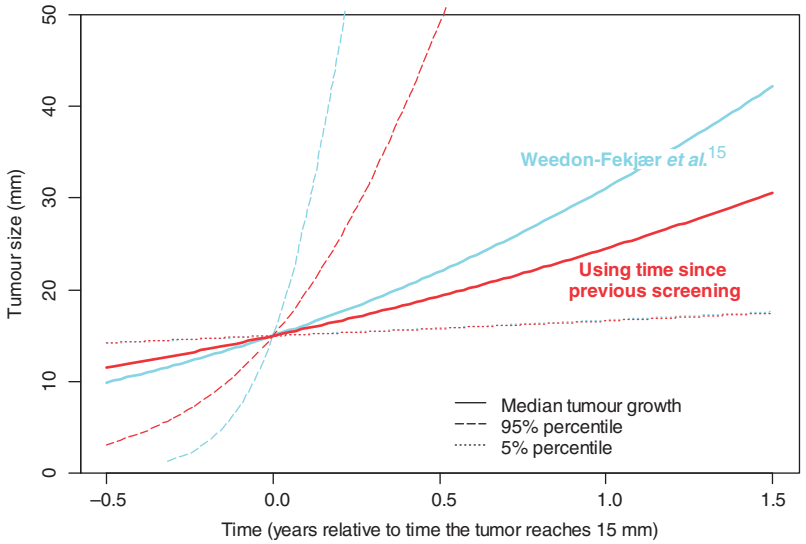

Figure 7 Tumour growth estimates using time since previous screening compared with previously published estimates based on interval cancers for women 50–69 years.[15] *Remark*: The 5% percentile for the estimate taken from Weedon-Fekjær *et al.*[15] is not seen in the figure, as it overlaps with the 5% percentile for the new estimate.

Table 4 Tumour growth and STS estimates using time since previous screening

	Main estimate	Estimate not corrected for increased HRT use
Time used from 10 to 20 mm (years)		
Mean	2.1 [1.8, 2.3]	2.7 [2.3, 2.9]
Standard deviation	2.2 [1.9, 2.4]	2.6 [2.3, 2.9]
Volume doubling time at 15 mm (days)		
25th percentile	76 [65, 88]	94 [79, 108]
50th percentile	152 [134, 170]	196 [171, 221]
75th percentile	303 [262, 337]	403 [334, 452]
Screening test sensitivity at		
5 mm	14 [12, 16]	13 [11, 14]
10 mm	58 [51, 66]	52 [46, 60]
15 mm	93 [89, 96]	89 [85, 94]

7 Discussion and recommendations for future studies

7.1 Model choice and design of study

With the strong correlation between MST and STS, and no clear biological definition of pre-clinical screening detectable state, interpretations of the basic Markov model are difficult. Hence, multi-state models or continuous tumour growth models often are a better choice if tumour growth measurements are available. The continuous tumour growth

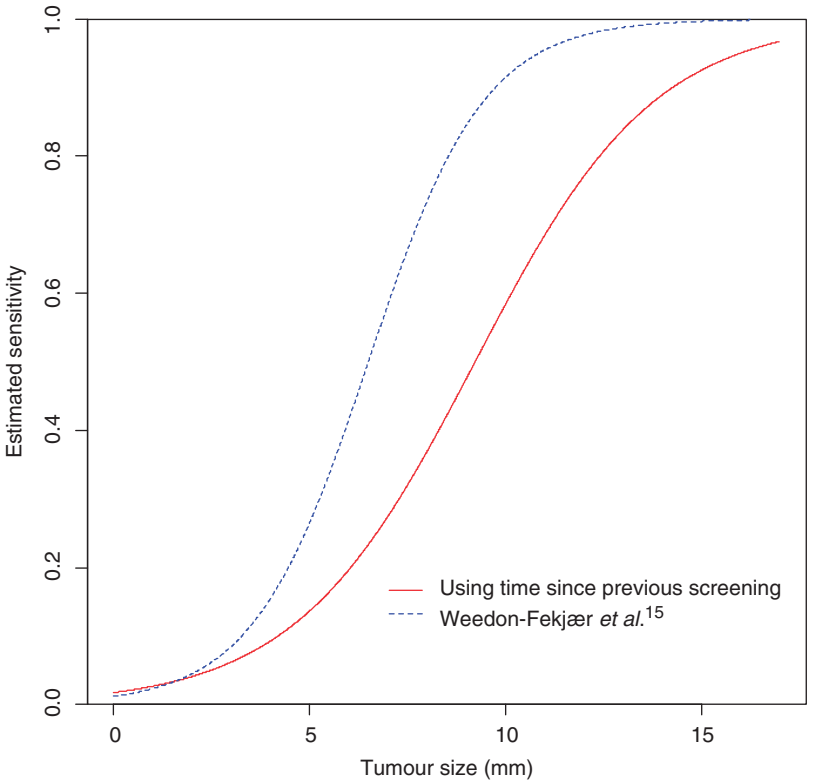

Figure 8 Screening test sensitivity estimates using time since previous screening compared with previously published estimates based on interval for women 50–69 years.[15]

models are somewhat more complex with more calculations/programming needed, but this is, for most applications, compensated by the clear link to biological measurements as tumour size. Hence, we would suggest that future studies carefully consider the newly suggested alternatives to the classical three-state Markov model of mammography screening. As seen in Weedon-Fekjær *et al.*[15] model fit for different models can be compared through cross-validation when evaluating different potential model choices. If only discrete stages in tumour development information are available, the multi-stage model is probably the best choice, while the continuous tumour growth model gives easy interpretable results when continuous tumour growth measurements are available.

Traditionally, estimation has been done by using data from the initial screening and the following interval. Modern screening programs have, however, often information about subsequent screening rounds and may have surveys of earlier mammography outside the screening program. As seen in Section 5, these data can also be utilised using either the classical Markov model and the continuous tumour growth model of Weedon-Fekjær *et al.*[15] Even more data is used in these study design, bootstrap confidence intervals using NBCSP data was wider for some estimates (Tables 2 and 3), likely as an effect of non-optimal model fit for the combined estimate. As we do not know which part of the model that best represent the true tumour progression process, we view the estimates

Table 5 Tumour growth and STS estimates with different STS for present and previous screening

Time used from 10 to 20 mm (years)	Estimate with bootstrap confidence interval	
Mean	1.8 [1.6, 1.9]	
Standard deviation	2.4 [0.8, 3.9]	
Volume doubling time at 15 mm (days)		
25th percentile	34 [0, 134]	
50th percentile	93 [0, 212]	
75th percentile	246 [195, 288]	
Screening test sensitivity at	First NBCSP appearance	Previous screening
5 mm	23 [19, 26]	26 [15, 37]
10 mm	88 [83, 93]	37 [25, 49]
15 mm	99 [99, 100]	49 [38, 62]

Remark: One of the hundred performed bootstrap replications were not included when calculating confidence intervals, due to non-convergence of the estimating procedural.

based on all available data as the most credible estimate. Hence, we would recommend future studies to acquire and use as much relevant data as possible when estimating tumour progression.

7.2 Tumour progression rates

Based on the Swedish Two County study, Tabar *et al.*[28] estimated MST to around 3.5 years and STS to 92–94%, both estimates increasing with age. When comparing the NBCSP MST and STS estimates seen here with Tabar *et al.*[28] and other Markov based studies,.[29] we find that MST is relatively long, while STS is relatively low. As discussed in articles by Weedon-Fekjær *et al.*[13,14] this could be an effect of more sensitive mammography due to increased quality, increasing the screening detectable phase and related MST, combined with lower STS for some woman due to increased HRT use. The true reasons for different MST and STS are, however, difficult to pinpoint with the somewhat unclear state definitions of the Markov model. Hence, comparisons to other screening programs using a continuous tumour growth model would be useful.

While the tumour progression estimates using the Markov model were hard to compare between studies, the continuously tumour growth model estimates have an easier interpretation with a direct link to tumour size. Using time since previous screening data, the large individual variation in tumour growth rates estimated from interval data is confirmed. Overall, STS estimates are lower than STS estimates found in Weedon-Fekjær *et al.*[15] and estimated tumour growth somewhat slower. The differences are, however, minor when using a model with different STS parameters for recent and previous screening. In contrast to the Markov model,[13] estimation with different present and former STS seem viable with reasonable stabile bootstrap estimates.

8 Final remarks

Both tumour progression and STS can be estimated indirectly, using population based mammography data. Traditionally, this has been done using Markov models, but

recently suggested models based on continuous tumour growth are likely to model the true cancer progression better. Combining the final estimates with other data, several key questions regarding screening can be illuminated. One example of such calculations is the work by the United States National Cancer Institute CISNET collaboration,[19] which combined data from several sources to estimate the attributable contributions of mammography and adjuvant therapy on breast cancer mortality in the United States from 1975 to 2000. As mammography screening has gone from randomised trials to regular screening programs, these modelling studies will be of increasing importance as randomised trials are no longer possible to conduct in populations with already wide spread screening.

Acknowledgements

Bootstrap replications in this paper have been produced using the Titan High Performance Computing facilities at the University of Oslo: http://hpc.uio.no/index.php/Titan.

References

1 Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. *Journal of Medical Screening* 2007; **14**: 174–7.
2 Spratt JA, von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. *Cancer* 1993; **71**: 2013–9.
3 Peer PG, van Dijck JA, Hendriks JH, Holland R, Verbeek AL. Age-dependent growth rate of primary breast cancer. *Cancer* 1993; **71**: 3547–51.
4 Spratt JS, Meyer JS, Spratt JA. Rates of growth of human solid neoplasms: Part I. *Journal of Surgical Oncology* 1995; **60**: 137–46.
5 Spratt JS, Meyer JS, Spratt JA. Rates of growth of human neoplasms: Part II. *Journal of Surgical Oncology* 1996; **61**: 68–83.
6 Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Computer Methods Programs Biomedicine* 1985; **20**: 79–93.
7 van Oortmarssen GJ, Habbema JD, van der Maas PJ, *et al*. A model for breast cancer screening. *Cancer* 1990; **66**: 1601–12.
8 Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984; **40**: 1–14.
9 Duffy SW, Chen HH, Prevost TC, Tabár L. Markov chain models of breast tumour progression and its arrest by screening. In Duffy SW, Hill C, Esteve J, eds. *Quantitative methods for the evaluation of cancer screening*. Arnold, London; 2001: 42–60.
10 Duffy SW, Chen HH, Tabar L, Day NE. 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* 1995; **14**: 1531–43.
11 Prevost TC, Launoy G, Duffy SW, Chen HH. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *American Journal of Epidemiology* 1998; **148**: 609–19.
12 World Health Organization. *IARC handbooks of cancer prevention: handbook 7: breast cancer screening*. IARC Press, Lyon, France; 2001.
13 Weedon-Fekjær H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Estimating mean sojourn time and screening sensitivity using questionnaire data on time since previous screening. *Journal of Medical Screening* 2008; **15**: 83–90.
14 Weedon-Fekjær H, Vatten LJ, Aalen OO, Lindqvist B, Tretli S. Estimating mean sojourn time and screening test sensitivity in

breast cancer mammography screening; new results. *Journal of Medical Screening* 2005; **12**: 172–8.
15 Weedon-Fekjær H, Lindqvist B, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Research* 2008; **10**: R41.
16 Shen Y, Zelen M. Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika* 1999; **86**: 503–15.
17 Cook RJ. A mixed model for two-state Markov processes under panel observation. *Biometrics* 1999; **55**: 915–20.
18 Berry DA, Cronin KA, Plevritis SK, *et al*. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine* 2005; **353**: 1784–92.
19 CISNET Home Page. CISNET webpage. Available at: URL: http://cisnet.cancer.gov/ (accessed 2007 March 23).
20 Wang H, Kåresen R, Hervik A, Thoresen SØ. Mammography screening in Norway; results from the first screening round in four counties and cost-effectiveness of a modeled nationwide screening. *Cancer Causes Control* 2001; **1**: 39–45.
21 Mammografiprogrammet (Norwegian Breast Cancer Screening Program). Available at: http://www.kreftregisteret.no/no/Forebyggende/Mammografiprogrammet/ 2009.
22 Cancer Registry of Norway. *Cancer in Norway 2006*. http://www.kreftregisteret no 2007 Available from: URL: http://www.kreftregisteret.no (accessed 2007 January 25).
23 Hartveit F, Maehle BO, Pettersen KC. Size of breast carcinomas at operation related to tumour growth rate. *Breast Cancer Research Treatise* 1987; **10**: 47–50.
24 Moller B, Weedon-Fekjær H, Hakulinen T, *et al*. The influence of mammographic screening on national trends in breast cancer incidence. *European Journal of Cancer Preview* 2005; **14**:117–28.
25 Bakken K, Alsaker E, Eggen AE, Lund E. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. *International Journal of Cancer* 2004; **112**: 130–4.
26 R core group. R. http://www r-project org 2003. Available at: URL: http://www.r-project.org (accessed 2006 September 11).
27 Weedon-Fekjær H. *Modelling breast cancer incidence, progression and screening test sensitivity using screening data*. University of Oslo, Norway; 2008.
28 Tabar L, Fagerberg G, Chen HH, *et al*. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* 1995; **75**: 2507–17.
29 Paci E, Duffy SW. Modelling the analysis of breast cancer screening programmes: sensitivity, lead time and predictive value in the Florence District Programme (1975–1986). *International Journal of Epidemiology* 1991; **20**: 852–8.

Appendix 1: Expected number of cases calculated for a basic Markov model of mammography screening

As explained in Duffy *et al*.[10] the expected number of cancer cases detected at screening and the following interval can be expressed as functions of STS, mean sojourn time and the expected breast cancer incidence without screening. Solving a special set of differential equations, so called Kolmogorov's equations, transition probabilities can be deduced and the probabilities for cancers at screening and the following interval calculated.[8] For a initial screening at age T, the probability of detecting a cancer at screening, $P(\lambda, STS)$, is given by:

$$P(\lambda, STS) = STS \cdot \left[\frac{r \cdot (e^{-\lambda \cdot T} - e^{-r \cdot T})/(r - \lambda)}{e^{-r \cdot T} + r\,(e^{-\lambda \cdot T} - e^{-r \cdot T})/(r - \lambda)} \right],$$

where *STS* is screening test sensitivity, $\lambda = 1/MST$, where *MST* = mean sojourn time, *r* the incidence of pre-clinical disease per time unit (typically 1 month).

As for interval cancers, the expected number of interval cancer cases in a short time unit, $\langle t_i - 1,\ t_i]$, is the sum of both cancer cases that have passed through the pre-clinical phase since last screening, and the overlooked cancer cases that have become clinical during the time interval. Mathematically, this can be expressed as:[10,11]

$$I\,(t_i, \lambda, STS) = r \cdot \left[1 - e^{-\lambda \cdot \left(t_i - \frac{1}{2}\right)}\right] + \frac{c \cdot (1 - STS)}{STS}\left[e^{-\lambda \cdot (t_i - 1)} - e^{-\lambda \cdot t}\right],$$

where *c* is the number of cancer cases that were detected at screening and *t* time from preceding screening examination. In practice, *c* can be estimated by the formulas above and the number of persons under observations in the given time interval.

As for repeated screening examinations, the probability of detecting cancer at age *T*, given a previous screening at age $T - X$ is:[13]

$$P\left(\begin{array}{c|c} \text{Tumor detectable} \\ \text{at screening} & \begin{array}{c}\text{A previous screening} \\ X \text{ years ago}\end{array}\end{array}\right)$$
$$= \frac{e^{-\lambda \cdot X - r \cdot (T - X)} - e^{-r \cdot T} + \left(e^{-\lambda \cdot T} - e^{-r \cdot (T - X) - \lambda \cdot X}\right) \cdot (1 - STS)}{e^{-\lambda \cdot X - r \cdot (T - X)} - \frac{\lambda \cdot e^{-J \cdot T}}{r} + \left(e^{-\lambda \cdot T} - e^{-r \cdot (T - X) - \lambda \cdot X}\right) \cdot (1 - STS)}$$

Appendix 2: R functions for Markov model

```
# Flexible functions for estimating mean sojourn time and screening sensitivity
# based on Markov models using both earlier screening and interval data
#
# Languane:    R (www.r-project.org)
# Licence:     GNU GPL version 2
# Written by:  Harald Weedon-Fekjær
# Url:         http://www.weedon-fekjaer.net/hwf/software/estMST/
# See also:    Upcoming article "Estimating screening test sensitivity and
#              tumour progression using tumour size and time since previous
#              screening", Weedon-Fekjaar, Tretli, Aalen
#              Statistical Methods in Medical Research (http://smm.sagepub.com )

# ----------------------------
# --- Input data: Example ---
# ----------------------------

scrdata <- data.frame(matrix(c(
  #scr, previous  , pyear     , obs ,
  1   ,NA         ,113238     ,972,
  1   ,0.66       ,36507      ,154,
  1   ,2          ,84449      ,405,
  1   ,4          ,38580      ,230,
  1   ,6.5        ,42091      ,289,
  0   ,0.125      ,160403     ,20,
  0   ,0.375      ,151232     ,64,
  0   ,0.625      ,146627     ,128,
  0   ,0.875      ,137044     ,142,
  0   ,1.125      ,128422     ,177,
  0   ,1.375      ,119225     ,166,
  0   ,1.625      ,110072     ,156,
  0   ,1.875      ,58457      ,75),byrow=T,ncol=4))
names(scrdata) <- c("scr","previous","pyear","obs")
   # Variables:
   # "screening" - Screening or interval data? (TRUE/FALSE)
   # "previous"  - Time since previous screening with NA or
   #               zero equally no previous screening
   # "pyear"     - Number of screened woman or number of person years for
   #               interval data
   # "obs"       - "Number of cancers detected" and
   #                 "Time since previous screening"

# ------------------------------------------------------------------------------
# - Function for estimating expected number of cases and estimating parameters -
# ------------------------------------------------------------------------------

estMST.expno <- function(data,J,ageT,lam,STS,intsize) {
  # Estimate expected number at screening given time since previous screening
  # via general functions of for transition probabilities
  # Without previous screening
  prob <- (J*(exp(-lam*ageT)-exp(-J*ageT))) /
          (J*exp(-lam*ageT)-lam*exp(-J*ageT))
  # With previous screening
  data$expno <- rep(NA,dim(data)[1])
  for (i in 1:dim(data)[1]) {
    if (data$scr[i]) {
      # Initial screening
      if (is.na(data$previous[i]) | data$previous[i]==0) {
        data$expno[i] <- prob*data$pyear[i]*STS
      } else { # Subsequent screening
        x <- data$previous[i]
```

```
        tmp <- exp(-lam*x-J*(ageT-x)) - exp(-J*ageT) +
               (exp(-lam*ageT)-exp(-J*(ageT-x)-lam*x))*(1-STS)
        prob.t <- tmp/ ( exp(-lam*x-J*(ageT-x)) - lam*exp(-J*ageT)/J +
                  (exp(-lam*ageT)-exp(-J*(ageT-x)-lam*x))*(1-STS))
        data$expno[i] <- prob.t*data$pyear[i]*STS
      }
    } else { #Interval
      x <- data$previous[i]
      c.scr <- prob*data$pyear[i]*(1/intsize)*STS
      term1 <- (J*data$pyear[i])*(1-exp(-lam*x))
      term2 <- ((1-STS)/STS) * (exp(-lam*(x-intsize/2))-exp(-lam*(x+intsize/2)))
      data$expno[i] <- term1 + term2*c.scr
    }
  }
  return(data)
}
# Function for estimating MST and STS given time since previous screening
estMST <- function(data,ageT,intsize=1/12,J=NULL) {
  # Find weighted square differences of observed-expected
  # (J="Expected incidence without screening", ageT="Age at screening")
  sdiff <- function(data,J,ageT,lam,STS) {
    exp.no   <- estMST.expno(data=data,intsize=intsize,J=J,ageT=ageT,
                             lam=lam,STS=STS)$expno
    exp.no   <- pmax(exp.no,0)
    w.u <- data$pyear*(exp.no/data$pyear)*(1-exp.no/data$pyear)
    w <- w.u/sum(w.u)
    return(sum( ((data$obs-exp.no)^2)/w ))
  }
  # Estimate:
  if (!is.null(J)) {
    opt.func <- function(x){
      return(sdiff(data=data,ageT=ageT,J=J,lam=1/x[1],STS=x[2]))
    }
    ret <- optim(c(1/4,.9),opt.func, lower=c(0,0), upper=c(Inf,1),
                 method="L-BFGS-B")$par
    names(ret) <- c("MST","STS")
  } else {
    opt.func <- function(x){
      return(sdiff(data=data,ageT=ageT,J=x[3],lam=1/x[1],STS=x[2]))
    }
    ret <- optim(c(1/4,.9,0.002),opt.func, lower=c(0,0),upper=c(Inf,1),
                 method="L-BFGS-B")$par
    names(ret) <- c("MST","STS","Bginc")
  }
  # Finishes:
  return(ret)
}

# ----------------------------
# --- Estimate parameters: ---
# ----------------------------

# Estimated MST and STS, with given breast cancer incidence without screening:
estMST(scrdata,ageT=60,intsize=1/4,J=0.00244)
 #*       MST       STS
 #* 4.6505864 0.7471805

# Estimated MST, STS and breast cancer incidence without screening:
estMST(scrdata,ageT=60,intsize=1/4)
 #*         MST         STS       Bginc
 #* 2.881840510 0.930649776 0.003182613
```