

Estimating mean sojourn time and screening sensitivity using questionnaire data on time since previous screening

Harald Weedon-Fekjær, Bo H. Lindqvist,
Odd O. Aalen, Lars J. Vatten and Steinar Tretli

Short title: Estimating sensitivity and sojourn time

Keywords: Breast cancer, screening, mean sojourn time, sensitivity, Markov model, opportunistic screening

Abbreviations: DCIS Ductal cancer in situ
HRT Hormone replacement treatment
MST Mean sojourn time
NBCSP Norwegian Breast Cancer Screening Program
STS Screening test sensitivity

Authors' affiliations:

Harald Weedon-Fekjær, Statistician, Department of Etiological Cancer Research, Cancer Registry of Norway, Norway and PhD student Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

Bo Lindqvist, professor, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

Lars J. Vatten, MD professor, Department of Community Medicine and General Practice, Norwegian University of Science and Technology, Trondheim, Norway

Odd O. Aalen, professor, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

Steinar Tretli, Research director, Cancer Registry of Norway, Norway and professor Department of Community Medicine and General Practice, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence to: Harald Weedon-Fekjær
Kreftregisteret, Institute of population-based cancer research
Montebello, N-0310 Oslo, Norway
E-mail: <harald.weedon-fekjaer@kreftregisteret.no>
Phone: +47 975 700 81

Abstract

Objectives: Mean sojourn time [MST] and screening test sensitivity [STS], is usually estimated by Markov models using incidence data from the first screening round and the interval between screening examinations. However, several screening programs do not have full registration of cancers submerging after screening, and increased use of opportunistic screening over time can raise questions regarding the quality of interval cancer registration.

Methods/Settings: Based on the earlier used Markov model, formulas for expected number of cases given time since former screening activity was developed. Using questionnaire data for 336 533 women in the Norwegian Breast Cancer Screening Program [NBCSP], mean square regression estimates of MST and STS were calculated.

Results: In contrast to the previously used method, the new approach gave satisfactory model fit. MST was estimated to 5.6 years for women aged 50-59 years, and 6.9 years for women aged 60-69 years, and STS was estimated to 55 and 60 percent, respectively. Attempts to add separate parameters for breast cancer incidence without screening, or previous STS, resulted in wide confidence intervals if estimated separately, and non-identifiably if combined.

Conclusion: Earlier NBCSP results (Weedon-Fekjaer at al. Journal of Medical Screening 2005; Vol. 12, No 4:172-178) of long mean sojourn time and low screen test sensitivity were confirmed with the new approach. Questionnaire data on time since previous screening can be used to estimate MST and STS, but the approach is sensitive to relaxing the assumptions regarding the expected breast cancer incidence without screening and constant screening test sensitivity over time.

Introduction

Breast cancer is difficult to prevent, but large-scale trials indicate that early treatment based on mammography screening is likely to reduce breast cancer mortality^{1,2}. As a result, mammography screening for early detection of breast cancer has become a standard health service in most developed nations. However, the optimal age range and time interval between screening rounds are still controversial subjects that need to be resolved.

Screening test sensitivity [STS] and the mean time in screening detectable phase, so-called mean sojourn time [MST], represent central parameters in the evaluation of screening procedures^{1,3}. Several methods are applied in the estimation of MST and STS, based on the increased incidence during screening and the reduced incidence after screening⁴⁻⁸. Most studies have used Markov models, where tumours are assumed to move from “no screening detectable breast cancer”, to “preclinical screening detectable breast cancer”, to “clinical detectable breast cancer”, at constant rates.

The common Markov approach is based on the observed incidence increase during screening and following incidence reduction. It is important to distinguish between cancers that are detected as a result of screening, and cancers that were clinically detected in the interval between screening examinations. The early randomized screening trials had reliable information about interval cancers, and the model gave acceptable fit to the observed data⁴⁻⁶. However, recently launched public screening programs are

accompanied by unregistered opportunistic screening that makes ascertainment of (true) interval cancer much less reliable, and a recent study shows substantial difficulties with model fit⁹.

Information on whether the breast cancers were diagnosed through routine screening or on the basis of clinical symptoms, is included in the Norwegian breast cancer report form. Hence, it should, at least in theory, be possible to identify cancers detected at opportunistic screening for women with no previous symptoms. However, while the general cancer registration in Norway is of very high quality, the indication for mammography is often insufficient and sometimes incorrect. The data should thus not be considered reliable in this respect. The Cancer Registry of Norway work towards mandatory reporting of screening at private institutes, but at present, reliable information regarding interval cancers is lacking. This as private mammography laboratories in practice perform both “routine opportunistic screening” and clinical mammography based on symptoms.

During the last decade, new screening procedures, improved screening technology¹⁰, and increased use of hormone replacement therapy¹¹ are factors that may have changed MST and STS substantially. Consequently, there is increasing uncertainty regarding the current validity of using previous MST and STS estimates. Since the previously used method may give biased estimates⁹, there is a need for more robust methods to estimate MST and STS. In addition, many screening programs do not include complete registration of interval cancers, which further emphasizes the need for additional methods of estimating

MST and STS that do not depend on information on interval cancers. Chen 2000¹² demonstrated that data from two screening rounds could be used to estimate MST and STS, and we will in this paper explore the possibility of using data from only one screening examination combined with questionnaire data on time since previous screening.

Material and methods

Data

In 1995, the Norwegian government initiated an organized population based screening program¹³, in which mammography results and interval cancer cases were carefully registered by the Cancer Registry of Norway. Initially the Norwegian Breast Cancer Screening Program [NBCSP] comprised four counties. Other counties were subsequently included, and by 2004 the screening program achieved nation wide coverage. All women between 50 and 69 years of age receive a written invitation every second year, and at first attendance, each woman is asked to answer a short questionnaire, including a question about time since previous screening (private or public). Until now, approximately 94 percent of the women have filled in the questionnaire.

The two-view mammograms are independently evaluated by two readers. The present study includes screening data from 1995 through 2002. Overall 78 percent of invited women participated during this period, resulting in 364 731 screened women 50-69 years of age. Among these women, 314 865 answered the question regarding former screening

experience at their first attendance in the NBCSP, and 113 238 reported no previous mammography. For an overview of the data used in the estimation, see Table 1.

Generally, there are good opportunities for population based cancer research in Norway. A high quality Cancer Registry¹⁴ and a unique personal identification number for each inhabitant in the country enable long-term follow-up of cancer incidence. Reporting of cancer is mandatory, and information is obtained independently from clinicians, pathologists and death reports.

To make results comparable with previous studies^{6,7}, we included all cases of ductal carcinoma in situ [DCIS] and all invasive breast cancer cases. Control estimations using only invasive cancer were done in order to study the effect of excluding DCIS. Only new primary breast cancers were included, and several tumours detected at the same time in one woman were counted as one case.

Estimation using time since previous screening

As for the earlier used method of estimating MST and STS, we assume a continuous time Markov model^{5,15}, consisting of three stages; “No screening detectable breast cancer”, “Preclinical screening detectable breast cancer” and “Clinical detectable breast cancer” (Figure 1). This corresponds to a model where women move from one stage to the next at a constant rate, with no possibility of regress once the tumour is visible at screening. The assumption of constant transition intensities imply a substantial variation in sojourn

times, and has given good model fit in several earlier mammography screening studies^{4,6-8}.

In practice, the proportion of positive mammograms will vary with time since previous screening, as some of the potential cancers will already have been detected at the previous screening. With increasing time since previous screening, the size of the reduction decreases, as fewer of the preclinical cancers were visible on the previous screening. With lower MST, the effect of the previous screening disappears faster, and we would suspect a steeper increase in the proportion of positive mammograms with time since previous screening. On the other hand, a high MST would be suspected to give a more gradual increase in the proportion of positive mammograms with time since previous screening, as the proportion of tumours visible on the last screening decreases more gradually. Similarly, a high STS would be expected to give few cancer cases among recently screened women, since most cancer cases in preclinical detectable phase then was suspected to already have been found at the recent screening.

Using screening data with time since previous screening, it is possible to utilize the differences in the proportion of positive mammograms between groups with different time since previous screenings to estimate MST and STS. To estimate the expected number of cases, we use the conditional probability of having a screening detectable cancer given the previous screening history. As deduced in the appendix I;

$$P\left(\begin{array}{c} \text{Tumor detectable} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{No} \\ \text{previous screening} \end{array}\right)$$

$$= \frac{J \cdot (e^{-\lambda t} - e^{-Jt})}{J \cdot e^{-\lambda t} - \lambda \cdot e^{-Jt}}$$
(1)

and

$$P\left(\begin{array}{c} \text{Tumor detectable} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{A previous screening} \\ x \text{ years ago} \end{array}\right)$$

$$= \frac{e^{-\lambda x - J(t-x)} - e^{-Jt} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)}{e^{-\lambda x - J(t-x)} - \frac{\lambda \cdot e^{-Jt}}{J} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)}$$
(2)

Where:

- STS is screen test sensitivity
- $\lambda = \frac{1}{\text{Mean sojourn time}} (= 1/MST)$
- J is the transition intensity from “No screening detectable cancer” to “Preclinical screening detectable cancer”; In practice equalling the expected cancer incidence without screening
- t the age of the screened woman

Since J is assumed known from other sources, we can estimate MST and STS using either non linear regression or maximum likelihood estimation. As MST and STS is known to vary with age¹, we have grouped our data into two age groups, estimating MST and STS separately for the two age groups. The data for each age group is given on the form

$$\{x_i, n_i, y_i\} \quad ; i = 1, \dots, I$$

where i is an index for screening history group, x are the times since previous screening, n are the corresponding number of screened woman, and y the corresponding number of cases found at screening.

Using this, the probability of detecting a tumour at screening becomes

$$P\left(\begin{array}{c} \text{Tumour detected} \\ \text{at screening} \end{array}\right) = \begin{cases} STS \cdot \frac{J \cdot (e^{-\lambda t} - e^{-Jt})}{J \cdot e^{-\lambda t} - \lambda \cdot e^{-Jt}} & \text{if the given woman has} \\ & \text{no earlier screening} \\ STS \cdot \frac{e^{-\lambda x - J(t-x)} - e^{-Jt} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)}{e^{-\lambda x - J(t-x)} - \frac{\lambda \cdot e^{-Jt}}{J} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)} & \text{if the given woman has} \\ & \text{been screened } x \text{ years ago} \end{cases} \quad (3)$$

and the expected number of cases, E_i , in screening history group i , is

$$E_i = n_i \cdot P\left(\begin{array}{c} \text{Tumour detected} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{Previous screening} \\ \text{history of group } i \end{array}\right) \quad (4)$$

Based on these formulas, MST and STS estimates are found by minimizing:

$$\sum_I^{i=1} \left(\frac{y_i - E_i}{w_i} \right)^2 \quad (5)$$

where y_i is the observed number of cases and w_i the weights in the non-linear regression. Using inverse variance weights, standardized to avoid estimation bias, the weights become:

$$w_i = \sqrt{\frac{n_i \cdot \frac{E_i}{n_i} \cdot \left(1 - \frac{E_i}{n_i}\right)}{\sum_I^{j=1} n_j \cdot \frac{E_j}{n_j} \cdot \left(1 - \frac{E_j}{n_j}\right)}} \quad (6)$$

Data handling

In the questionnaire, time since previous 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 (i.e. the values of x_i defined above). Other choices for representative points were also tested, to make sure that this choice did not substantially influence the results (data not shown).

NBCSP offers screening to all women in the defined population, so no parallel control group is available to estimate incidence without screening. In addition, commitment to screening can, and probably does vary by individual risk factors, so women not attending are not considered a suitable control group. Hence, background incidence was calculated from historical data combined with an estimated time trend. In practice, incidence data from 1990-1994 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.¹⁶. 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 for the previously used method that utilized the incidence of interval cancers, the assumed background incidence without screening (J) is essential in the estimation. Hence, our estimated background incidence is important, but the strong increase in hormone replacement therapy use in the 1990s¹⁷ may have influenced and distorted the historical time trends. In a large Norwegian cohort study, Bakken et al. 2004¹⁷ found a relative risk of 2.1 of breast cancer for current use of hormone replacement therapy. Combining sales figures with risk estimates, Bakken et. al. 2004 have estimated that 27 percent of breast cancer cases could be attributable to the use of hormone replacement therapy among Norwegian women aged 45-64 years. The use of hormone replacement therapy increased sharply from our historical comparison data (1990-1994) to the estimation period (1996-2002), and this could have changed the historical trends significantly. Combining hormone replacement therapy sales figures and risk estimates found in Bakken et. al. 2004, breast cancer incidence is estimated to increase 21 percent from our historical data to our estimation period due to increased HRT use. Hence, estimates were (when otherwise not noted) calculated with a 21 percent added background incidence, to account for a probable increase in hormone replacement therapy use.

One possible problem when using time since previous screening in the analysis is potential bias caused by different frequencies of both opportunistic/private screening and breast cancer according to socioeconomic status. To evaluate the probable impact of this potential problem, adjusted estimates were calculated taking into account the observed differences in hormone replacement therapy use (ever), children (yes/no), number of

children, age at first birth, alcohol consumption, education and county of residence, among woman with and without a previous screening history. Calculations were done in a three step approach; first, the relationship between covariates and the risk of breast cancer among women with no reported previous screening history was estimated, using logistic regression. Then an adjusted risk of breast cancer for women without previous screening history was calculated, combining the estimated covariate coefficients with the distribution of covariates among women with and without a previous screening history. Finally, the new adjusted level of breast cancer cases among woman with no reported screening history was used to calculate adjusted MST and STS estimates.

In order to compare the new estimation method with the previously used method, estimates from Weedon-Fekjær et al. (2005) have been included in Table 2.

Estimation of model parameters

Parameters are estimated using non-linear weighted mean square regression. The main estimates of MST and STS were given with robust 95 percent bias corrected bootstrap confidence intervals, calculated from 2000 bootstrap replications. For a validation of the model fit, observed versus expected values were plotted. All calculations, simulations and figures were performed using the R statistical package¹⁸.

Results

MST was estimated to 5.6 years for women aged 50-59 years, and 6.9 years for women aged 60-69 years (Table 2), and STS was estimated to 55 and 60 percent, respectively.

Comparing the new estimates with estimates based on the number of interval cancers, MST estimates were 1.7 years higher for woman 50-59 years, and 1.9 years higher for woman 60-69 years, with both confidence intervals overlapping (slightly) with results from the previously used method. As for STS, there were substantial differences, with STS for woman 50-59 years of 55 % vs. 75 % with the previously used method based on interval cancers, and for woman 60-69 years 60 % vs. 85 %, respectively.

The earlier estimation method, using number of interval cancers, gave only moderate fit to the observed NBCSP data⁹, while the new method using time since previous screening fitted well to the observed data (Figure 2).

Without the assumed 21 percent added background incidence due to increased hormone replacement therapy use (Table 3), MST estimates increased from 5.6 to 7.2 years for women aged 50-59 years, and from 6.9 years to 8.6 years for women aged 60-69 years. The corresponding STS decreased to 52 percent and 58 percent for the two age groups. Excluding DCIS cases, MST was estimated to 4.7 years for woman 50-59 years of age, and 6.0 years for women in the age group 60-69 years, while STS estimates remained constant. Adjusting for different risk factors among women with and without previous opportunistic screening had little effect on the estimated values (Table 3).

Discussion

Most previous studies of MST and STS have used the Markov model approach of estimation by number of cancer cases at first attendance and in the interval between

screening examinations^{1,7,19}. However, given the recent increase in opportunistic screening between ordinary screening rounds, the usefulness of the model may be questioned, and Weedon-Fekjær et al. (2005) found only moderate fit to the observed NBCSP data. To solve the problems caused by unregistered opportunistic screening, a new approach using time since previous screening is suggested. Applied to the NBCSP dataset, the model gave good fit to the observed data (Figure 2), and the method seems to be a good supplement for the previously used method. The new approach requires good quality questionnaire data about time since previous screening, but allows MST and STS estimation in screening programs without registration of interval cancer cases. As in Chen 2000¹² there is no need of information regarding interval cancer, and in addition only data from one screening examination can be used in populations with widespread non registered screening.

Based on the previously used method and the Swedish Two County study data¹, Tabar et al. (1995) estimated MST to around 3.5 years and STS to 92-94 percent, both estimates increasing with age. Compared to the Swedish Two County study and several other studies^{1,4,19,20,20}, Weedon-Fekjær et al. (2005) found significantly higher MST and lower STS in the Norwegian Breast Cancer screening program [NBCSP]. When adjusting for a higher incidence due to increased HRT use, MST in the NBCSP program was estimated to 3.9 years for women aged 50-59 years, and 5.0 years for women aged 60-69, and the corresponding STS was estimated to 75 and 85 percent, respectively. The authors suggested that some of these differences could be attributed to bias due to opportunistic

screening, but the new estimating method presented here confirms the long MST and low STS, without the problem of bias from opportunistic screening.

As MST and STS often are highly correlated in the simulations estimation of these Markov models⁷, it could be suggested that a higher STS and lower MST could explain the observed data. Locking STS to 90 %, MST decreases towards values observed in earlier studies, but with a considerable loss of model fit (Figure 3).

However, the new method could also be subject to bias, especially by variations in STS over time, as one assumption of the new method is constant STS over time. If the sensitivity of the screening test increases, the estimated STS for the new test will be too low. Usually, bias due to variation in STS would only constitute a minor problem in breast cancer mammography screening, since STS changes slowly over time, but new screening methods and routines could in some situations bias the estimates. One example is digital mammography, but use of digital mammograms in Norway was limited during the study period. A related problem is possible systematic differences in STS between the official screening program and the previous screening tests conducted outside the program. In Norway, this may explain the lower STS estimates seen with the new method, since a substantial part of previous mammograms were conducted by private practitioners with a possibly different set of routines than the public screening program.

Generally, with STS varying from previous to present screening, STS estimates using time since last screening can be seen as a compromise between previous and present STS.

On the other hand, STS estimates based on the number of interval cancers only estimate the present STS. As for MST, the estimate is in practice probably mostly based on the observed incidence at screening, combined with STS and the expected incidence in the absence of screening. With lower STS estimates, MST estimates will probably also be biased (upward). Hence, the approach using time since last screening should mainly be used when there are no indications of considerable changes in STS over time.

The probably most problematic assumption is the estimated breast cancer incidence without screening. Removing the correction for a likely increase in breast cancer incidence due to more hormone replacement therapy use, MST increased considerably while STS increased slightly. Hence, the longer MST and low STS could not be an effect of an over correction. Still, the effect of HRT could be even higher than assumed, but it is not likely to explain all the observed differences from earlier screening programs/trials.

As there is some uncertainties regarding both the assumed constant STS and the assumed background incidence without screening, two model extensions taking into account these possible departures were tested. As shown in appendix II, a change in STS from the previous tests is impossible to separate from a change in the assumed background incidence without screening. Hence, the details of these calculations must be interpreted with great care. However, the long MST and low STS of both models further confirm a possible change in MST and STS compared to earlier screening programs/trials (Table 4).

Another possible bias could come from non-progressive (or regressive) cancers. Earlier studies of MST and STS have assumed only progressive cancers⁴⁻⁸, and good model fit

indicates that this is a minor problem with the new method. There are few reports of regression of invasive breast cancers^{21,22}, but there is uncertainty regarding non-invasive cancer (DCIS). Some studies have suggested that only a small proportion of DCIS regresses²³, but this question is currently under discussion. To explore the effect of possible regressive DCIS cases, special calculations excluding DCIS were performed. As shown in Table 3, the effects of excluding DCIS on the estimated MST and STS were quite modest, yielding a slightly shorter MST and minor changes in sensitivity.

Since opportunistic screening is not uniformly distributed among different social groups, variations in the probability for earlier screening may correlate with different risk factors. It is conceivable that this could have biased our new estimates. However, adjustment for socioeconomic factors had only a negligible effect on the estimates (Table 3).

The increased MST implies that patients spend more time in the preclinical screening detectable phase, and may indicate an increase in screening efficiency. During the last years, substantial effort has been made to increase the quality^{10,24} of mammographic images and other aspects of mammography screening. These improvements have most likely resulted in earlier detection of tumours, and preliminary findings indicate that the NBCSP is successful in detecting small tumours. Compared to EU guidelines, the NBCSP performs well on tumour size, the proportion of detected DCIS cases, and the proportion of lymph node positive cases²⁵. Comparing NBCSP with the Nijmegen and Utrecht trials²⁰, the proportion of DCIS in the first screening round was 17 percent and 13 percent, respectively. Similarly, the proportion of invasive cancer cases 20 mm or

larger in the first round was 19 percent for NBCSP, compared to 28 percent combined for the Nijmegen and Utrecht trials.

The low sensitivity may seem strange in the light of a possible increase in mammographic quality. However, one should note that sensitivity in these Markov models is not an absolute measure of quality, and that the expected number of diagnosed cancer cases at each screening round was actually higher than reported in most previous studies (illustrated by $MST \times STS$). Therefore, the new techniques may enable the detection of additional small tumours, but simultaneously yield low screening test sensitivity. Use of hormone replacement therapy is also known to reduce STS^{11,26,27}, at least partly by increasing breast density. In Norway, the use of hormone replacement therapy was common during the study period¹⁷, and the reduced STS may, to a certain extent, be an effect of hormone replacement therapy, especially among woman 50-59 years of age, in whom HRT use was highest²⁸.

To summarize, the new method using time since previous screening seems to be a good substitute for the previously used method of estimation based on number of interval cancer cases. The new method is robust against unregistered opportunistic screening, and gives good model fit to mammography screening data, but it is not recommendable to extend the model with additional parameters for either a change in STS over time or breast cancer incidence without screening. It requires information on time since last screening, but it does allow for estimation of MST and STS in screening programs without registration of interval cancers. These analyses confirm estimates from

Weedon-Fekjær et al. 2005, with higher MST and lower STS in the NBCSP compared to results from other mammography programs or trials. New mammography methods have probably increased the number of tumours that may be detected at screening, while increasing use of hormone replacement therapy has probably reduced STS. This study is (probably) the first to apply time since previous screening in the estimation of MST and STS, and we therefore encourage new studies on data from other programs to verify our results. An implementation of the new method for the free R statistical package¹⁸ is available at <http://folk.uio.no/hfe/software/estMSTprev/>.

Acknowledgements and financial support:

The authors would like to thank Bjørn Møller, Solveig Hofvind, Unn E. Hestvik and Tom Grotmol for valuable comments on the article, and Solveig Hofvind and Wenche Melbye for sharing their excellent knowledge about the Norwegian breast cancer screening programs and database. The project has received financial support from the Norwegian Cancer Society (grant E03031/001), and is connected to the Centre for Biostatistical Modelling in the Medical Sciences group at University of Oslo, Norway.

References

1. 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.
2. World Health Organization. IARC Handbooks of Cancer Prevention: Handbook 7: Breast Cancer Screening. IARC Press, 2001.
3. Boer R, de Koning HJ, van der Maas PJ. A longer breast carcinoma screening interval for women age older than 65 years? *Cancer* 1999;**86**:1506-10.
4. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984;**40**:1-14.
5. Duffy SW, Chen HH, Prevost TC, Tabár L. Markov chain models of breast tumour progression and its arrest by screening. *Quantitative Methods for the Evaluation of Cancer Screening*. Arnold, 2001:42-60.
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. *Statistics in Medicine* 1995;**14**:1531-43.
7. Prevost TC, Launoy G, Duffy SW et al. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *American Journal of Epidemiology* 1998;**148**:609-19.
8. Habbema JD, van Oortmarssen GJ, Lubbe JT et al. The MISCAN simulation program for the evaluation of screening for disease. *Computer Methods and Programs in Biomedicine* 1985;**20**:79-93.

9. Weedon-Fekjaer H, Vatten LJ, Aalen OO et al. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening; new results. *Journal of Medical Screening* 2005;**12**:172-8.
10. Haus AG. Historical technical developments in mammography. *Technol.Cancer Res.Treat.* 2002;**1**:119-26.
11. Litherland JC, Stallard S, Hole D et al. The effect of hormone replacement therapy on the sensitivity of screening mammograms. *Clinical Radiology* 1999;**54**:285-8.
12. Chen HH, Kuo HS, Yen MF et al. Estimation of Sojourn Time in Chronic Disease Screening without Data on Interval Cancers. *Biometrics* 2000;**56**:167-71.
13. Wang H, Kåresen R, Hervik A et al. Mammography screening in Norway; results from the first screening round in four counties and cost-effectiveness of a modeled nationwide screening. *Cancer Causes and Control* 2001;**1**:39-45.
14. Cancer Registry of Norway. Cancer in Norway 2004. <http://www.kreftregisteret.no>
15. Chen HH, Duffy SW, Tabár L. A Markov Chain Method to Estimate the Tumour Progression Rate from Preclinical to Clinical Phase, Sensitivity and Positive Predictive Value for Mammography in Breast Cancer Screening. *The Statistician* 1996;**45**:307-17.
16. Moller B, Weedon-Fekjaer H, Hakulinen T et al. The influence of mammographic screening on national trends in breast cancer incidence. *European Journal of Cancer Prevention* 2005;**14**:117-28.
17. Bakken K, Alsaker E, Eggen AE et al. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. *International Journal of Cancer* 2004;**112**:130-4.
18. R core group. R. <http://www.r-project.org>

19. 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.
20. van Oortmarssen GJ, Habbema JD, van der Maas PJ et al. A model for breast cancer screening. *Cancer* 1990;**66**:1601-12.
21. Ross MB, Buzdar AU, Hortobagyi GN et al. Spontaneous regression of breast carcinoma: follow-up report and literature review. *Journal of Surgical Oncology* 1982;**19**:22-4.
22. Powles TJ, Hickish T. Breast cancer response to hormone replacement therapy withdrawal. *Lancet* 1995;**345**:1442.
23. Yen MF, Tabar L, Vitak B et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *European Journal of Cancer* 2003;**39**:1746-54.
24. Perry N, Broeders M, de Wolf C, Törnberg S. European Guidelines for Quality Assurance in Mammography Screening, 3 ed. ed. European Commission, 2001.
25. Hofvind S, Wang H, Thoresen S. Do the results of the process indicators in the Norwegian Breast Cancer Screening Program predict future mortality reduction from breast cancer? *Acta Oncologica* 2004;**43**:467-73.
26. Porter PL, El Bastawissi AY, Mandelson MT et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *Journal of the National Cancer Institute* 1999;**91**:2020-8.
27. Kavanagh AM, Mitchell H, Giles GG. Hormone replacement therapy and accuracy of mammographic screening. *Lancet* 2000;**355**:270-4.

28. Bakken K, Eggen AE, Lund E. Side-effects of hormone replacement therapy and influence on pattern of use among women aged 45-64 years. The Norwegian Women and Cancer (NOWAC) study 1997. *Acta Obstetricia et Gynecologica Scandinavica* 2004;**83**:850-6.
29. Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* 1986;**42**:855-65.

APPENDIX I: The probability of a screening detectable cancer given previous screening history

The basic Markov model illustrated in Figure 1 has intensity transition matrix

$$Q = \begin{pmatrix} -J & J & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & 0 \end{pmatrix}$$

where $\lambda = \frac{1}{\text{Mean sojourn time}}$ and J , the transition intensity from the state “No screening detectable cancer” to “Preclinical screening detectable cancer”, in practice equals the expected cancer incidence without screening per time unit (typical one year).

For a given woman, let $X(t)$ be the state (0, 1 or 2) at age t , and define:

$$P_{ij}(t) = P(X(s+t) = j | X(s) = i) \quad \text{for all } t > 0, s > 0$$

Applying Kolmogorov’s equations, we deduce $P(t)$ through standard differential equations and find that²⁹:

$$P(t) = \begin{pmatrix} e^{-Jt} & \frac{J \cdot (e^{-\lambda t} - e^{-Jt})}{J - \lambda} & 1 - \frac{J \cdot e^{-\lambda t} - \lambda e^{-Jt}}{J - \lambda} \\ 0 & e^{-\lambda t} & 1 - e^{-\lambda t} \\ 0 & 0 & 1 \end{pmatrix} \quad (7)$$

Applying this, we get that:

$$\begin{aligned}
 & P\left(\begin{array}{c|c} \text{Tumor detectable} & \text{No} \\ \text{at screening} & \text{previous screening} \end{array}\right) \\
 &= P(X(t)=1 | X(t) < 2) = \frac{P_{01}(t)}{P_{00}(t) + P_{01}(t)} \\
 &= \frac{J \cdot (e^{-\lambda t} - e^{-Jt})}{J \cdot e^{-\lambda t} - \lambda \cdot e^{-Jt}}
 \end{aligned} \tag{8}$$

While this probability has been given in earlier articles^{6,29}, we will in this paper need the corresponding probability giving a previous screening at age $t - x$. Defining V as the event “The tumour was not detected at the previous screening”, the conditional probability can be written as:

$$\begin{aligned}
 & P\left(\begin{array}{c|c} \text{Tumor detectable} & \text{A previous screening} \\ \text{at screening} & x \text{ years ago} \end{array}\right) \\
 &= P(X(t)=1 | \{X(t-x)=0 \cup (X(t-x)=1 \cap V=0)\} \cap X(t) < 2) \\
 &= \frac{P(X(t)=1 \cap X(t-x)=0 \cap X(t) < 2) + P(X(t)=1 \cap X(t-x)=1 \cap V=0 \cap X(t) < 2)}{P(X(t-x)=0 \cap X(t) < 2) + P(X(t-x)=1 \cap V=0 \cap X(t) < 2)} \\
 &= \frac{P_{00}(t-x)P_{01}(x) + P_{01}(t-x)P_{11}(x)(1-STS)}{P_{00}(t-x)(P_{00}(x) + P_{01}(x)) + P_{01}(t-x)P_{11}(x)(1-STS)}
 \end{aligned} \tag{9}$$

$$= \frac{e^{-\lambda x - J(t-x)} - e^{-Jt} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)}{e^{-\lambda x - J(t-x)} - \frac{\lambda \cdot e^{-Jt}}{J} + (e^{-\lambda t} - e^{-J(t-x) - \lambda x}) \cdot (1 - STS)}$$

An intuitive approximation

As an alternative to the precise probability calculations above, a good approximation can also be deduced using a slightly more intuitive approach. As each woman has a probability of J of developing a screening detectable cancer each year, and each cancer has a mean time in preclinical screening detectable phase of $\frac{1}{\lambda}$, we get that:

$$P\left(\begin{array}{c|c} \text{Tumor detectable} & \text{No} \\ \text{at screening} & \text{previous screening} \end{array}\right) \approx J \cdot \frac{1}{\lambda} \quad (10)$$

Looking at the precise formula presented earlier (8), this is the limit as $t \rightarrow \infty$ using that $J < \lambda$.

For a woman with previous screening, this probability can be adjusted by subtracting the probability of having had a tumour found at previous screening that still would have been in preclinical screening detectable phase without the previous screening. Using the Markov property, a portion $e^{-\lambda x}$ of all cancer cases found at screening x years earlier would be expected to have remained in preclinical screening detectable phase. Hence, the probability

of having a screening detectable tumour given a previous screening x years ago can be approximated by:

$$\begin{aligned}
 & \text{Probability of having a cancer in preclinical screening} \\
 & \text{detectable stage without previous screening} \\
 P \left(\begin{array}{c} \text{Tumor detectable} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{A previous screening} \\ x \text{ years ago} \end{array} \right) & \approx \overbrace{J \cdot \frac{1}{\lambda}}^{\text{Probability of having a cancer in preclinical screening}} - \underbrace{\left(J \cdot \frac{1}{\lambda} \right) \cdot STS \cdot e^{-\lambda \cdot X}}_{\substack{\text{Probability of having had a breast cancer} \\ \text{that was found the previous screening}}} \underbrace{e^{-\lambda \cdot X}}_{\substack{\text{Probability of the tumour still} \\ \text{being in preclinical screening} \\ \text{detectable stage after } x \text{ years} \\ \text{given no screening examination}}} \quad (11) \\
 & = J \cdot \frac{1}{\lambda} \cdot (1 - STS \cdot e^{-\lambda \cdot X})
 \end{aligned}$$

Looking at the precise formula presented earlier (9), this is a good approximation to the $t \rightarrow \infty$ limit, since in practice $J \ll \lambda$.

A recalculation of the numbers found in Table 2 and Table 3, using this approximation, shows that MST estimates at the most changes 0.1 years, while STS estimates are within +/- 1 percentage point (data not shown).

APPENDIX 2: Relation between models with additional parameters for STS of previous screening or a new parameter for the underlying breast cancer incidence without screening

From equation (11) and (4) we have that:

$$P\left(\begin{array}{c} \text{Tumor detectable} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{A previous screening} \\ x \text{ years ago} \end{array}\right) \approx J \cdot \frac{1}{\lambda} \cdot (1 - STS \cdot e^{-\lambda \cdot x})$$

and

$$E_i = STS \cdot n_i \cdot P\left(\begin{array}{c} \text{Tumor detectable} \\ \text{at screening} \end{array} \middle| \begin{array}{c} \text{A previous screening} \\ X_i \text{ years ago} \end{array}\right)$$

Defining STS_1 as screening test sensitivity of the previous screening, and STS_2 as

screening test of the present screening, and $\tau = \frac{STS_2}{STS_1}$ we find that:

$$\begin{aligned} E_i &\approx STS_2 \cdot n_i \cdot J \cdot \frac{1}{\lambda} \cdot (1 - STS_1 \cdot e^{-\lambda \cdot X}) \\ &= [STS_1 \cdot \tau] \cdot n_i \cdot J \cdot \frac{1}{\lambda} \cdot (1 - STS_1 \cdot e^{-\lambda \cdot X}) \\ &= STS_1 \cdot n_i \cdot [J \cdot \tau] \cdot \frac{1}{\lambda} \cdot (1 - STS_1 \cdot e^{-\lambda \cdot X}) \end{aligned}$$

Hence, we have shown that the ratio $\tau = \frac{STS_2}{STS_1}$, between STS_1 and STS_2 , has the same effect

on the (approximated) expected values as a multiplicative shift in J . In practice this

implies that a model with separate STS for previous and present screenings, and a model

with an added parameter for the underlying breast cancer incidence without screening, will

have the same effect on the estimated MST and STS_1 .

Tables:

Table 1: Basic data used in the estimation;

Screening outcome as function of time since previous screening

	50-59 years		60-69 years	
	Woman (n)	Cases (n)	Woman (n)	Cases (n)
	<hr/>		<hr/>	
No previous Screening	68 436	469	44 802	503
Time since previous screening				
0-1 years	25 103	95	11 404	59
1-3 years	57 726	245	26 723	160
3-5 years	25 020	134	13 560	96
> 5 years	26 022	171	16 069	118
Expected incidence without screening	190 per 100 000 person year		219 per 100 000 person year	

Table 2: Mean sojourn time [MST] and screening test sensitivity [STS] estimates; Using time since previous screening (new method), or number of cancers found at screening combined with the interval cancer rate (*previous method*)

	Estimates using no of cancers as function of time since previous screening <i>(new method)</i>		Estimates using no of cancers at screening and interval cancer rates* <i>(previous method)</i>	
	50-59 years	60-69 years	50-59 years	60-69 years
MST (years)	5.6 [4.0– 6.6]**	6.9 [5.5 – 7.8]**	3.9 [3.2 – 4.2]***	5.0 [4.3–5.5]***
STS (percent)	55 [43 – 67]**	60 [49 – 71]**	75 [70 – 82]***	85 [80 – 90]***

* Excluding earlier screened women from screening data

** 95% bias corrected bootstrap confidence interval

*** 95% smoothed bias corrected bootstrap confidence interval

Table 3: Mean sojourn time [MST] and screening test sensitivity [STS] estimates based on different assumptions; Main estimates, excluding Ductal cancer in situ [DCIS], without correcting the assumed background incidence due to increased hormone therapy [HRT] use, and adjusting for different covariates among women with and without previous screening.

Assumptions	MST estimate (years)		STS estimate (percent)	
	50-59 years	60-69 years	50-59 years	60-69 years
	Basic estimate	5.6	6.9	55
Excluding DCIS	4.7	6.0	55	60
Without correcting for increased HRT use	7.2	8.6	52	58
Adjusting for different risk factors [*]	5.5	6.9	61	59

^{*} Hormone therapy use (ever), children (yes/no), number of children, age at first birth, alcohol consumption, education and county of residence

Table 4: Estimates (with 95% bias corrected bootstrap confidence interval) of different models with additional parameters for either screening test sensitivity [STS] of previous screening or underlying breast cancer incidence without screening

	Estimates	
	50-59 years	60-69 years
Common for both models		
MST (years)	4.2 [0.8– 5.8]	11.9 [0.0 – 18.5]
STS (percent)	57 [42 – 71]	56 [43 – 67]
Model specific parameters		
STS present screening	87 [74 – 100]	43 [0 – 73]
Relative background incidence*	1.52 [0.80 – 2.07]	0.77 [0.13 – 1.33]

* Compared with historical trend

Figures:

Figure 1: Markov model for breast cancer screening

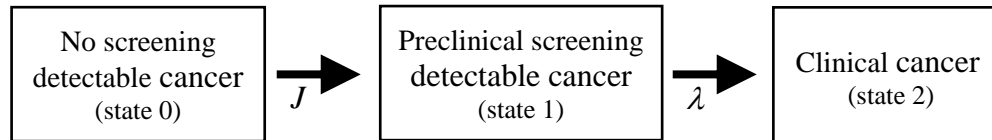


Figure 2: Model fit using time since last screening

(dotted lines mark 95% confidence interval for observed values)

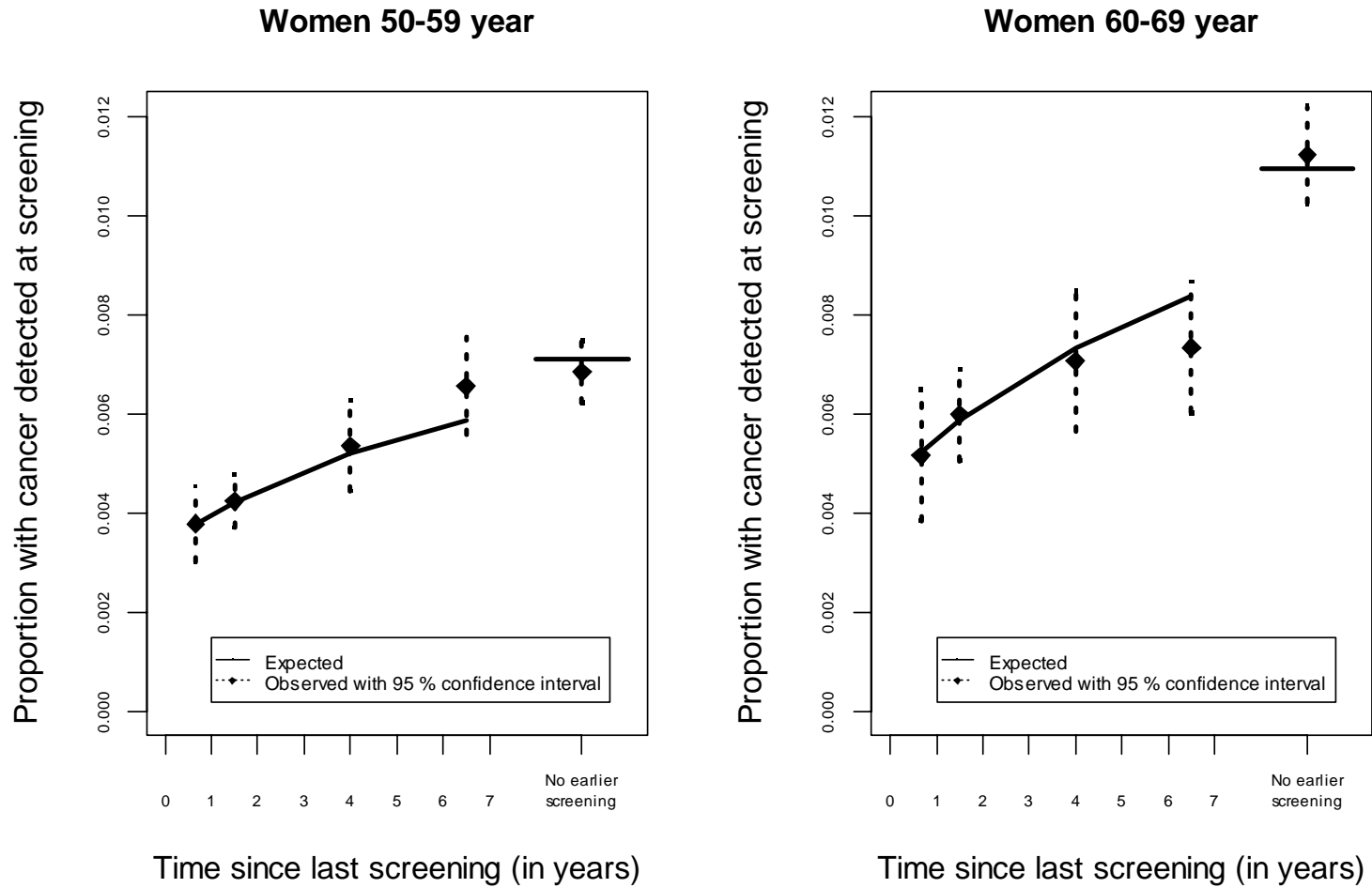


Figure 3: Model assuming screening test sensitivity [STS] of 90 % vs. separate STS variable

(MST estimate with STS=0.9 is 3.0 for woman 50-59 years, and 4.8 for woman 60-69 years.)

