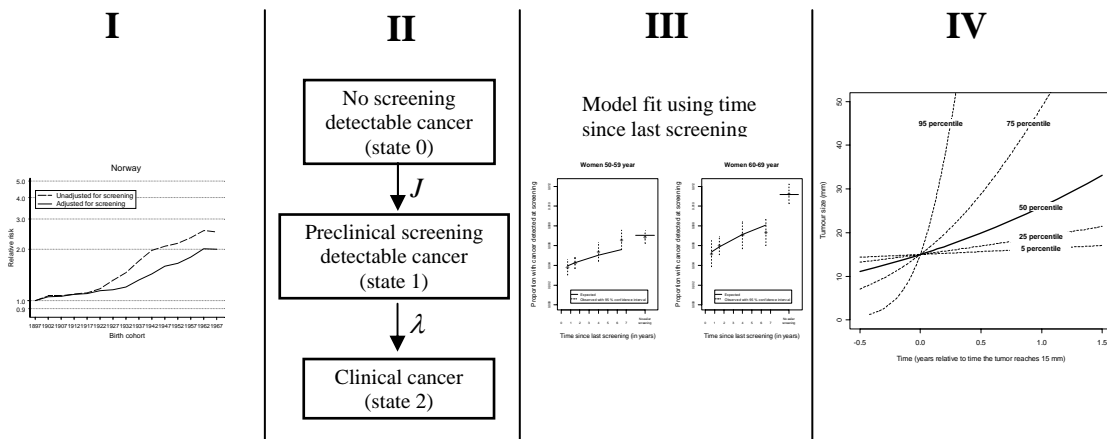
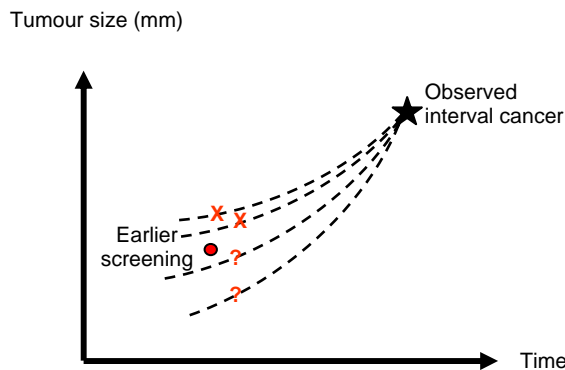




Modelling breast cancer incidence, progression and screening test sensitivity using screening data

PhD summary for Harald Weedon-Fekjær

University of Oslo, Department of biostatistics, July 2007



© Harald Weedon-Fekjær, 2007

(except published papers where the copyrights have been transferred to the journal/publisher)

ISBN 978-82-8072-255-3

Errata:

- In the legends for figure 13, the time interval has been corrected from "1990-2006" to "1987-2006" after the committee evaluated the thesis.

Contact address of author:

Harald Weedon-Fekjær (*statistician*)
E-mail: haraldwf@weedon-fekjaer.net
Phone: +47 97570081 (private mobile)
Webpage: <http://www.weedon-fekjaer.net/hwf/>

Supervisors:

Steinar Tretli Research director (statistician), Cancer Registry of Norway
Professor, Norwegian University of Science and Technology

Co-supervisors:

Lars J. Vatten Professor (medical doctor), Norwegian University of Science and Technology
Bo H. Lindqvist Professor (statistician), Norwegian University of Science and Technology
Odd O. Aalen Professor (statistician), University of Oslo

Adjudicative committee:

Professor J. Dik F. Habbema, Erasmus MC, Netherlands
Professor Giske Ursin, University of Southern California, USA
Researcher Randi M. Selmer, The Norwegian Institute of Public Health, Norway

Financial support:

Norwegian Cancer Society and the Cancer Registry of Norway

Table of content:

Terms and abbreviations	4
List of papers.....	5
Acknowledgements.....	6
1 General introduction.....	7
1.1 Breast cancer — A serious health problem.....	7
1.2 Risk factors and primary prevention.....	9
1.3 Symptoms, classification and treatment of breast cancer	10
1.4 Cancer growth and development.....	11
1.5 Mammography screening for early detection of breast cancer	13
2 Aims of the thesis	17
3 Materials and methods.....	19
3.1 The Cancer Registry of Norway	19
3.2 The Norwegian Breast Cancer Screening Program	19
3.3 Combining data from different sources.....	21
3.4 Estimating the effect of breast cancer screening on incidence rates.....	22
3.5 Estimating breast cancer progression: the Markov model	23
3.6 Estimating breast cancer growth.....	28
3.7 Estimation in practice: challenges and solutions	29
4 Summary of results.....	31
4.1 Summary of paper I.....	31
4.2 Summary of paper II	32
4.3 Summary of paper III.....	33
4.4 Summary of paper IV.....	34
5 Discussion: methodological choices and potential biases	35
5.1 Bias caused by variations in HRT use	35
5.2 Bias from unregistered screening.....	38
5.3 MST and STS in the NBCSP	41
5.4 Comparing the Markov model and the new growth model (of paper IV)	46
6 Conclusions	49
7 References	51
8 Appendix: Quantifying the level of unregistered mammography screening.....	57
9 Papers (I-IV).....	61
10 Summary for the general public (in Norwegian)	Backside of thesis

Terms and abbreviations

CI	Confidence interval
DCIS	Ductal carcinoma in situ A non-invasive breast tumour limited inside the milk ducts of the breast
IARC	International Agency for Research on Cancer: WHO's cancer research institute
HRT	Hormone replacement therapy
MST	Mean sojourn time See sojourn time
ML	Maximum likelihood A common estimating technique (1)
NBCSP	Norwegian Breast Cancer Screening Program
NCI	National Cancer Institute The United State's publicly founded national cancer institute
NLSR	Non-linear mean square regression A common estimating technique (1)
Over-diagnosis	Women who would not have had any breast cancer diagnosis in their life time without participating in screening (2)
Opportunistic screening	Screening outside an official program without prior clinical symptoms
Sojourn time	Time in screening detectable phase before clinical detection
STS	Screening test sensitivity
WHO	World Health Organisation The health unit of United Nations

List of papers

- I. Møller, B., **Fekjaer, H.**, Hakulinen, T., Tryggvadottir, L., Storm, HH., Talback, M. and Haldorsen, T. (2005); *“The influence of mammographic screening on national trends in breast cancer incidence”*, European Journal of Cancer Prevention; 14(2):117-128

- II. **Weedon-Fekjær, H.**, Vatten, L. J., Aalen O. O., Lindqvist, B., Tretli, S. (2005); *“Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening; New results”*, Journal of Medical Screening, Vol 12, no 4: 172-178

- III. **Weedon-Fekjær, H.** , Lindqvist, B., Aalen O. O., Vatten, L. J., Tretli, S. *“Estimating sojourn time and screening sensitivity using questionnaire data on time since previous screening”* (submitted)

- IV. **Weedon-Fekjær, H.**, Vatten, L. J., Lindqvist, B. , Aalen O. O., Tretli, S. *“Breast cancer tumour growth estimated through mammography screening data”* (submitted)

Acknowledgements

To the women of my life; Maja (daughter), Susanne (wife), Turid (mother) and Silje (sister),
with hope of better knowledge of a killer disease

Finishing this thesis, I am extremely grateful to my supervisor Steinar Tretli, and look forward to future co-operation at the Cancer Registry of Norway. Steinar has always been available, interested in my work, and supported me one hundred percent. Without his great experience and superb advices, I would never have finished within financing without large frustrations and sacrificing important time with my family. I am indeed lucky to have had such a supervisor!

In addition to my main supervisor, I want to express my gratitude to my co-supervisors. Lars J. Vatten has carefully gone through all my manuscripts, and taken my scientific English and presentation to a new level. Bo H. Lindqvist has double-checked all my suggested mathematical theories, and given superb proposals. Odd Aalen has come up with many theoretical suggestions, and I greatly enjoyed mixing with his BMMS group at Oslo University.

The work has been carried out at the Cancer Registry of Norway, and my sincerest thanks go to director Frøydis Landmark and administrative chief Egil Engen for building a great cancer registry and always giving me the best working conditions. Next, I wish to thank the Norwegian Cancer Society for financing the project, and giving me the possibility to apply highly interesting mathematical-statistical theory on such an important field. It is with veneration I receive the financial support, and I hope the work some day can lead to better management of breast cancer and breast cancer screening.

I am also grateful to the co-authors of the first paper, Timo Hakulinen, Tor Haldorsen, Laufey Tryggvadóttir, Hans H. Storm, Mats Talbäck and especially my good friend Bjørn Møller, for working out a highly cited paper that became the start of my PhD. Many of my colleagues at the Cancer Registry of Norway have read different manuscripts, and I am grateful for their help and for making the Cancer Registry a great place to work. I especially want to thank Unn E. Hestvik, Trude Eid Robsahm, Tom Grotmol, Tom K. Grimsrud, Tom B. Johannesen, Astri Syse and Hans-Olav Adami. With little previous knowledge of breast cancer, Solveig Hofvind, Steinar Thoresen and Wenche Melbye's explanations regarding breast cancer and the screening database have been invaluable.

The last weeks have been intense with work from morning to afternoon, but with superb support from my wife Susanne. In addition, I want to thank my father Hans Olav, my mother Turid and my parents in law, Pomme and Harald, for helping us looking after our two lovely children, ensuring that they never had too long days in the preschool despite my need to make progress in this work. Look back on the years working with this thesis, it has been a hectic time with both Maja and Magnus arriving during the PhD work, but it has also given me the greatest moments of life.



Harald Weedon-Fekjær, Oslo, Norway (July 2007)

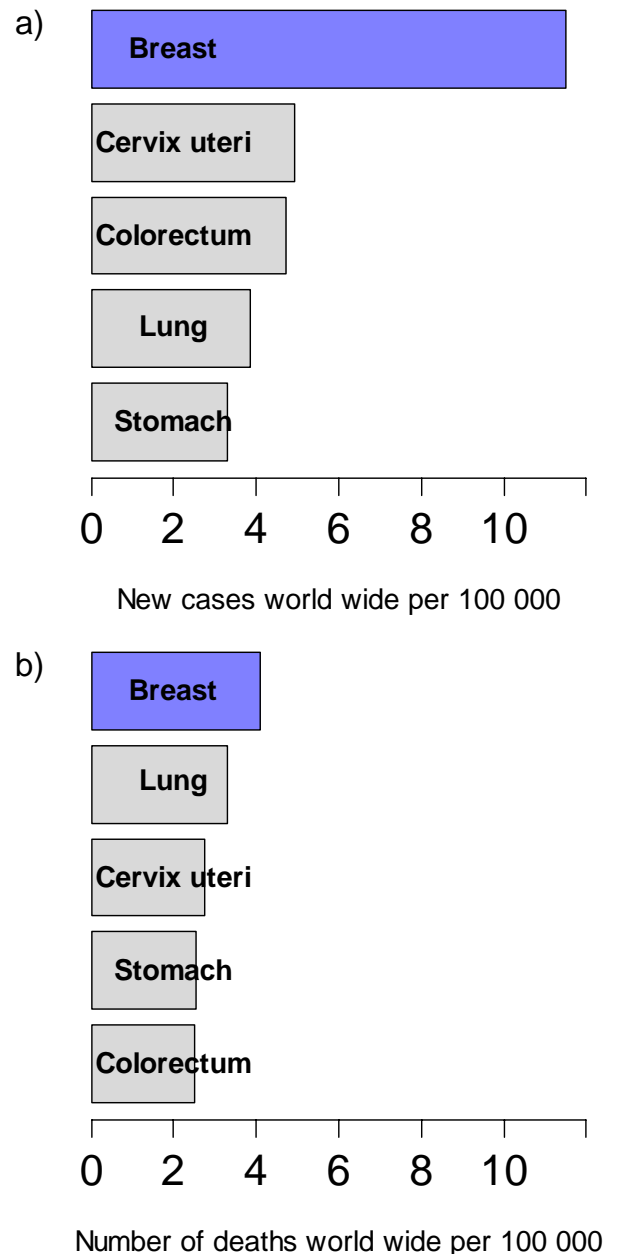
1 General introduction

1.1 Breast cancer — A serious health problem

Breast cancer is the most common cancer among women world wide (4). Each year, more than one million new cases of breast cancer are diagnosed (Figure 1), and over four hundred thousand deaths are recorded (5). In addition, breast cancer incidence increases in nearly all countries (6;7), and the disease has one of the youngest median age at diagnosis of the most common cancers (8).

The incidence of breast cancer is much lower in Asia, Africa and South America than in North America, Australia and Europe (Figure 2), with age standardized incidence rates varying from 18 per 100 000 women in Eastern Asia to 90 per 100 000 women in North America (4). Even though breast cancer incidence rates today are highest in Europe and North America, migrant studies (9) and risk factors closely connected to standard of living (age at menarche, number of children etc.), indicate that breast cancer may in the future become a larger problem in Asia and the rest of the world (9-11).

Figure 1: The five most common cancers among women world wide in a) number of cases and b) number of deaths



Data: "Global Cancer Statistics, 2002" (3)

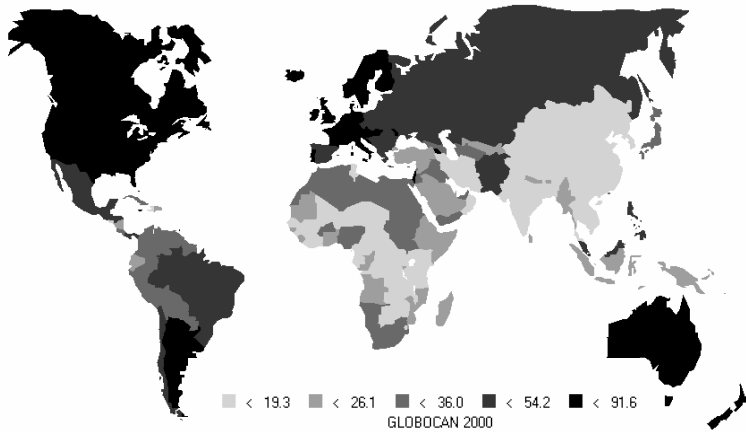


Figure 2:
World wide incidence of breast cancer
in year 2000

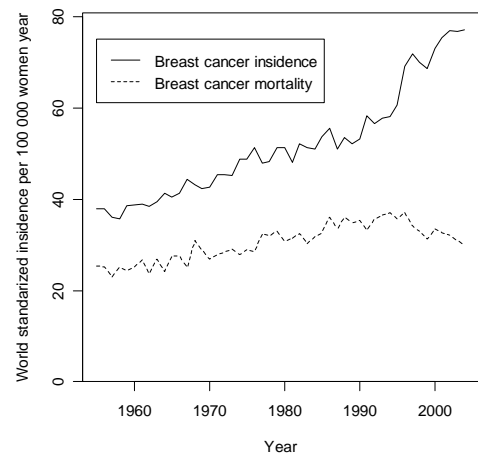
Source: Figure is generated by
Globocan 2000. (12)

Breast cancer risk in Norway

As in most west European countries, Norway has a relatively high incidence of breast cancer. There has been a continuous increase in breast cancer incidence since the start of Norwegian cancer registration in 1953, with an accelerating increase in the 1990s largely due to the introduction of mammography screening (Figure 3). Some of the increase may be an effect of earlier diagnosis strategies, indicated by the increasing proportion of lower staged cancers (13). However, the increase of lower staged cancers has not been followed by a fully compensating fall in higher staged cancers, indicating an overall increase in breast cancer frequency. There has been an increase in the survival rate of Norwegian breast cancer patients, but due to the high incidence there are still a large number of breast cancer deaths (Figure 3). With the high living standard and good health services of Nordic countries, most life threatening diseases occur at high age, but breast cancer also frequently occur in middle aged women, resulting in breast

cancer being the most important cause of lost life years for Norwegian women under 65 years (14). Norwegian women today have an estimated breast cancer lifetime risk of 10.8 %¹, and breast cancer accounts for 3.3 % of the deaths among Norwegian women.

Figure 3: Breast cancer incidence and mortality for Norwegian women 1955-2004



Data: Cancer Registry of Norway

¹ Lifetime risk was calculated using 2005 numbers from Statistics Norway and the Cancer Registry of Norway. The estimates may be somewhat exaggerated since official mammography screening still is in an introduction phase in some counties, with especially high breast cancer incidence.

1.2 Risk factors and primary prevention

The considerable increase in breast cancer risk in developed countries emphasizes the wish for an effective primary prevention through reduction in known risk factors. There has been an extensive research on breast cancer during the last decades², and the knowledge about risk factors has increased substantially. Well confirmed risk factors for breast cancer are high age at first birth, few children, early menarche, late menopause, a family history of breast cancer (genetic predisposition), a previous benign breast disease, high oestrogen levels including the use of hormone replacement therapy [HRT], elevated body height, high postmenopausal or low pre-menopausal body mass index, high breast density, and exposure to ionizing radiation (15). In addition, there are several probable risk factors, e.g. high alcohol consumption, low physical activity and short lactation period, for which more research is needed.

While few women receive substantial levels of ionizing radiation, other risk factors such as early menarche, high age at first birth, and few children contribute considerably to the overall risk. In Italy, the effect of parity patterns was estimated to contribute to 38 % of the observed breast cancer cases, compared with a scenario where all women were giving birth before reaching 20 years of age (16). Hence, with the exception of HRT use (17), most breast cancer risk factors are difficult to modify and not suited for public health programs (18).

² Pubmed (18) includes over eight thousand articles with the word “breast cancer” published in 2006, and Google (<http://www.google.com>) reports 34 million hits on “breast cancer research” (April 2007).

1.3 Symptoms, classification and treatment of breast cancer

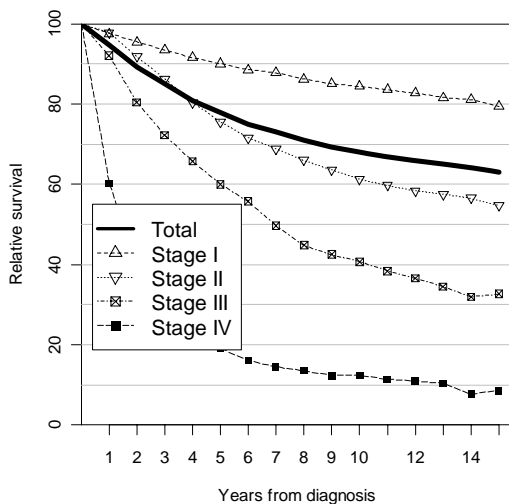
The most common clinical symptom of breast cancer is a palpable tumour. Following the suspicion of breast cancer, the diagnosis is established by clinical examination, mammography and fine needle aspiration (FNA) biopsy with cytological examination. If the diagnosis is still unclear, an open biopsy is performed.

Breast cancers are classified in four stages from small cancers (<20 mm in diameter) with no spread to lymph nodes, to metastatic cancers spreading beyond the breast and invading lymph nodes. Prognosis is closely related to stage, with far better prognosis for lower staged breast cancers (Figure 4). As

seen in Figure 5, prognosis has increased steadily for all stages during the last decades.

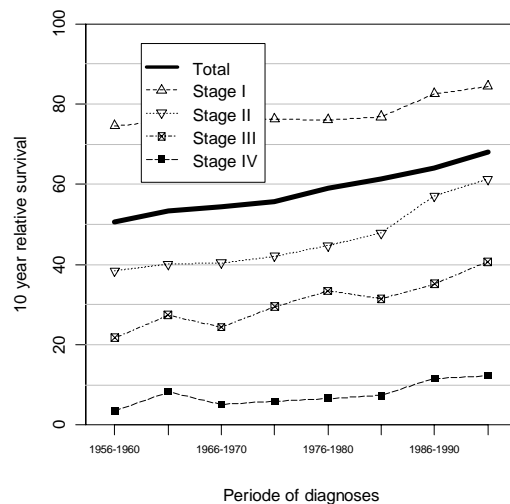
As for most cancers, the main treatment is surgical removal of the tumour and surrounding tissue, combined with post-operative radiation and chemotherapy. Earlier, removal of the whole breast and related lymph nodes was the standard treatment (mastectomy). As a result, many women had extensive side effects such as swelling caused by excess fluid build-up after lymph node removal (lymphedema). After several studies showing little or no increased mortality with breast-conserving surgery, it became increasingly common during the

Figure 4: 15 year survival for (female) breast cancer patients diagnosed before the official screening program (1991-1995), illustrating the relatively long period of excess mortality related to breast cancer



Data: Cancer Registry of Norway

Figure 5: Trends in 10 year breast cancer survival among Norwegian women diagnosed between 1956 and 1995. As breast cancer patients have a relatively long period of excess mortality, hence 10 year survival was chosen instead of the standard 5 years



Data: Cancer Registry of Norway

1990s. Recently (2002), the positive results were confirmed in a long-term follow up study (19), and breast-conserving surgery has now become the standard treatment of care.

After introduction of breast-conserving surgery, breast cancer surgery has recently developed further. Radioactive blue dye is today injected around the tumour site, and observed on its path to the lymph node

1.4 Cancer growth and development

Generally, cancer can be seen as uncontrolled cell growth that do not respect the barriers of neighbouring tissues and organs (22). With genetically mutated cells, the tumour grows and develops with a higher growth rate than the surrounding cells. At one point the tumour may invade neighbouring tissues, defined as invasive cancer. Once the cells have gained malignant characteristics, tumour progression is, without treatment, in most cases probably an irreversible process.

The typical clinical growth rates of tumours are, however, difficult to quantify as most cancers are surgically removed shortly after detection. Breast cancer is a very heterogenic disease with many different types of genetic alterations (23). As a result, clinical studies have shown large variations in tumour growth rates (24). Several animal studies and studies using cell cultures have assessed the growth

draining fluid from the cancer. Hence, the sentinel lymph node that the cancer is most likely to have metastasized to is located, making it possible to remove fewer lymph nodes and further reduce the side effects of surgery (20). To further improve the prognosis, adjuvant use of the anti-oestrogen drug tamoxifen has become widespread, improving the prognosis considerably for oestrogen receptor positive tumours (21).

of mammary cancer cells (25;26). These studies provide valuable information concerning different promoters of tumour growth, but the relevance for estimating the actual growth rates of clinical human tumours is questionable. Observational studies of patients have examined tumours that were initially overlooked on earlier mammograms, or tumours in women who refused treatment (27-30). These studies have confirmed large variations in cancer growth, but the studies are typically small and probably influenced by length time bias, since slow-growing tumours are relatively long time in pre-clinical stages that are visible on mammograms. This limits their potential use, increasing the need for further studies (31).

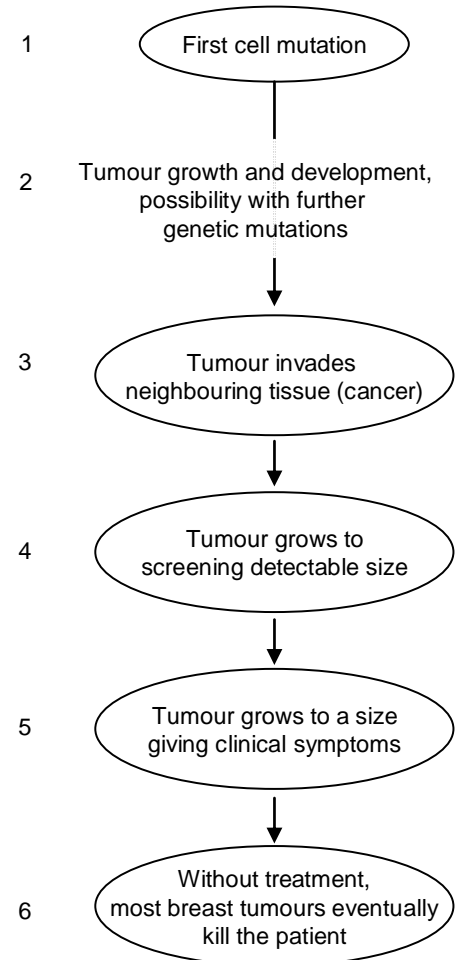
If a malignant tumour is not removed, it will in most cases progress and eventually kill the patient. With its removal, most breast cancer

patients survive. Still, patients have a risk of the tumour relapsing even 10-15 years after surgery, as it is difficult to ensure that every cancer cell is removed. In addition to local recurrence, metastases to other organs are common, accounting for a large proportion of breast cancer deaths. Hence, it is important to diagnose and remove the cancer before it metastasizes. The actual time of metastasis of breast tumours is highly debated, and probably varies considerably between tumours.

The relation between tumour development and risk factors is complicated (32), but some risk factors probably act as initiators and others as promoters. One example is hormone therapy used by women to limit menopausal nuisances. Studies have shown an increased risk of breast cancer even shortly after the start of hormone therapy use, and a corresponding normalization of risk rather soon after discontinued use. The short time span between the start of medication and increase risk, probably does not allow cancers initiated after the start of medication to reach clinical detection by the time risk starts to increase. Hence, hormone therapy probably acts as a promoter of already initiated tumour processes.

Even with increasing biological knowledge, important questions remain, such as time of metastasis and good estimates of clinical growth rates. Combined with the effect of screening programs, these are important issues for future studies.

Figure 6: Typical breast cancer development. Individual variations are, however, large, with some tumours never becoming visible on mammograms, while others become visible already as non-invasive DCIS. A key point related to screening is the timing of metastases. If metastases occur before the cancer reaches screening detectable size, the potential gain of screening is limited

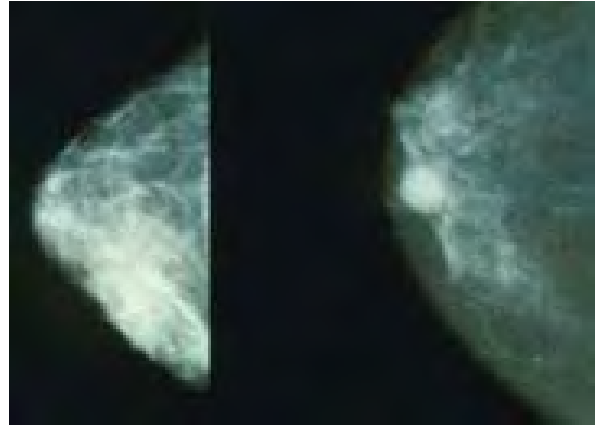


1.5 Mammography screening for early detection of breast cancer

With an increasing incidence of breast cancer, few possibilities for primary prevention, and substantial mortality even after decades with advances in treatment, medical doctors have been looking for new ways to combat the large number of breast cancer deaths. As survival is substantially better for tumours diagnosed in early stages (Figure 4), a natural strategy would be to advance the time of diagnosis (33;34).

Although breast self examination is a natural starting point for earlier diagnosis, randomized trials with organized instruction in breast self examination have shown little or no effect in reducing the number of breast cancer deaths (35;36). Hence, focus has turned to more advanced diagnostic techniques. Mammography (Figure 7) is a technique capable of guiding the final diagnosis of clinical breast cancers, exploiting the different absorption of X-rays of different tissues (37). In the 1960s and 1970s, several large randomized clinical trials using mammography as a screening test for pre-clinical cancers were initiated (38-41), with the goal of reducing the number of breast cancer deaths through earlier treatment. After 13 years of follow-up, the Swedish Two-County Trial of women 40-74 years of age reported a 30 % reduction in breast cancer mortality in the screening vs. control arm of

Figure 7: Examples of mammograms: normal (left) and cancerous (right)



Source: http://en.wikipedia.org/wiki/Breast_cancer/

the study (42). The effects were smaller in several other trials, but most reported considerable reductions in breast cancer mortality (38;43). Largely based on the randomized trials, the World Health Organisation [WHO] recommended mammography as a routine health service for middle aged women (44).

Following positive reports from the randomized trials, many countries initiated organized mammography screening. Later mammography screening has been highly debated (45-47), and both the WHO's International Agency for Research on Cancer [IARC] and the National Cancer Institute [NCI] in the United States initiated working groups to re-evaluate the evidence of a mortality reduction following mammography screening. Both working groups concluded in

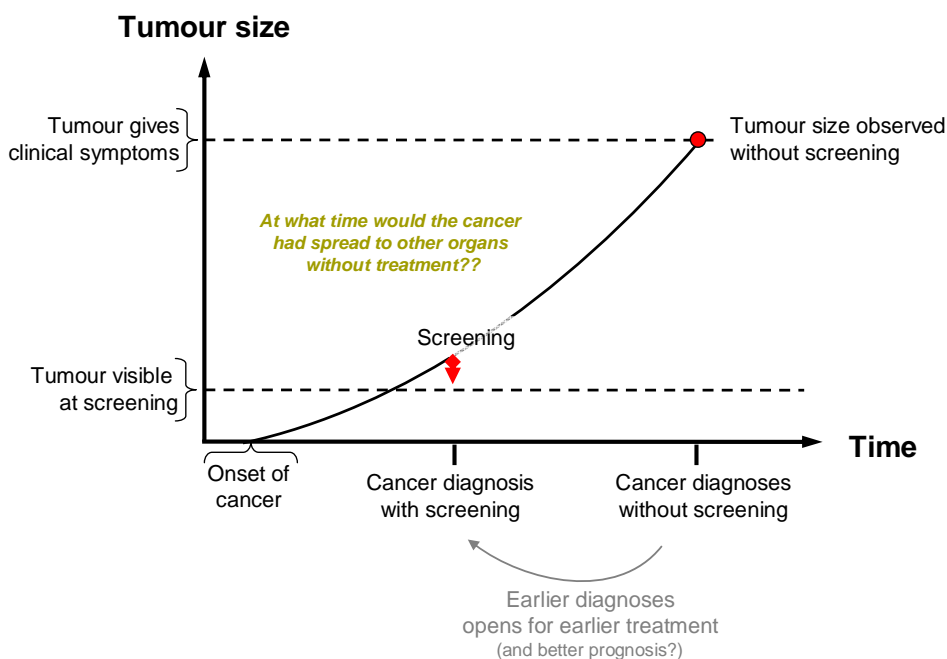
2002 that there was sufficient evidence of a mortality reduction following early treatment based on mammography screening (48;49).

While the debate regarding the pros and cons of mammography screening has continued, the official recommendations of regular mammography screening of middle aged women have not been changed. At the same time, there is a continued interest in the analyses of large screening programs in order to improve the quality and recommendations of mammography screening (50).

The basic principle of mammography screening as a health service is earlier treatment and better prognosis through earlier

diagnosis (Figure 8). Even though most researchers believe earlier detection and treatment to a certain degree increases survival, there are large uncertainties regarding the critical time of diagnosis/treatment (51). As many breast cancer deaths are due to metastases to other parts of the body, a vital question is when these metastases occur, and whether mammography screening can move the time of diagnosis sufficiently so that the cancer can be treated before it metastasizes. In addition, a screening method which successfully moves the time of diagnosis, will also sometimes detect cases in woman that without screening would die of other causes shortly after screening. Hence, a certain degree of over-

Figure 8: The idea behind mammography screening: advance the diagnosis and the related treatment to a time before the cancer has spread to other organs



diagnosis is inevitable (for definition, see page 4), and must be balanced against the gain of the given screening method. Both the gain of screening and the level of over-diagnosis depends on the screening method used, the implementation, the frequency (typically 1-3 year), and the age groups screened. In practice, screening recommendations differ considerable between the United States (NCI) and WHO (IARC), both regarding screening frequency and recommended age span. Screening older women (> 70 years of age) with a high rate of non-breast cancer related deaths, may result in

a relatively high level of over-diagnosis, while screening young women with low breast cancer risk may result in very few detected cases. In addition, mammography as a screening method is probably less cost-effective in young women (< 50 years of age) with dense pre-menopausal breasts and a probable higher mean tumour growth rate. Clinical and laboratory studies can provide information regarding some of these questions, but as nearly all cancers are treated, the full natural breast tumour progression may not be observed. Hence, many questions still remain, and more research is warranted.

2 Aims of the thesis

- 1) Estimate the impact of mammography screening on breast cancer incidence, and estimate screening corrected incidence trends in the Nordic countries (paper I)**
- 2) Utilize mammography screening related variations seen in breast cancer incidence, to estimate breast cancer progression, growth, and screening test sensitivity (paper II-IV)**
- 3) Improve and develop new, statistical methods to estimate breast cancer progression, growth, and screening test sensitivity based on variations in breast cancer incidence caused by mammography screening (paper III & IV)**

When developing methods for predicting future cancer incidence in the Nordic countries (8;52), we needed a way of correcting for past screening information, and utilizing future screening plans, to improve the predictions. Hence, the idea of paper I was conceived, studying breast cancer incidence trends in relation to official screening patterns in the Nordic countries. As a result, mammography screenings' effect on breast cancer incidence and the underlying trends in breast cancer risk corrected for screening in the Nordic countries were estimated.

With the high quality data from the Cancer Registry of Norway and the observed effect of mammography screening on breast cancer incidence trends, we wanted to utilize this to improve the knowledge about cancer progression and screening test sensitivity. This resulted in three papers: paper II applying an earlier used model, paper III adjusting the model of paper II to a different dataset avoiding problems with unregistered opportunistic screening, and paper IV developing a new model utilizing tumour measurements to improve the estimates of cancer progression.

3 Materials and methods

3.1 The Cancer Registry of Norway

The Cancer registry of Norway has covered the Norwegian population with complete registration since 1953 (12). Reporting of cancer cases is mandatory, and information is obtained independently from clinicians, pathologists and death certificates (see Figure 9). The registration is known for high quality data, indicated by the low number of cases reported based on death certificates only. In 2004, only 0.3 % of reported breast cancer cases were based on death certificate only.

From 1960 onwards, every inhabitant of Norway has been assigned a unique personal identification number, used by the Norwegian Population Registry at Statistics Norway, the Cancer Registry of Norway, and the Norwegian Breast Cancer Screening Program. This enables complete follow-up over time, and the possibility of linking data from several sources precisely.

3.2 The Norwegian Breast Cancer Screening Program

In 1995, the Norwegian government initiated a population based screening program (53) administered by the Cancer Registry of Norway. The goal was a 30 % reduction in breast cancer deaths through earlier treatment. A large number of process indicators are carefully registered in the Norwegian Breast Cancer screening database. Initially, the Norwegian Breast Cancer Screening Program [NBCSP] included four counties. Other counties were subsequently included, and in

2004 the screening program achieved nation wide coverage (Figure 10). Every second year all women between 50 and 69 years of age receive a written invitation with suggested time and place for mammography screening, and a possible reminder after 1-4 months. The two-view mammograms from participating women are independently evaluated by two readers.

Figure 9: Sources of information and the registration process at the Cancer Registry of Norway

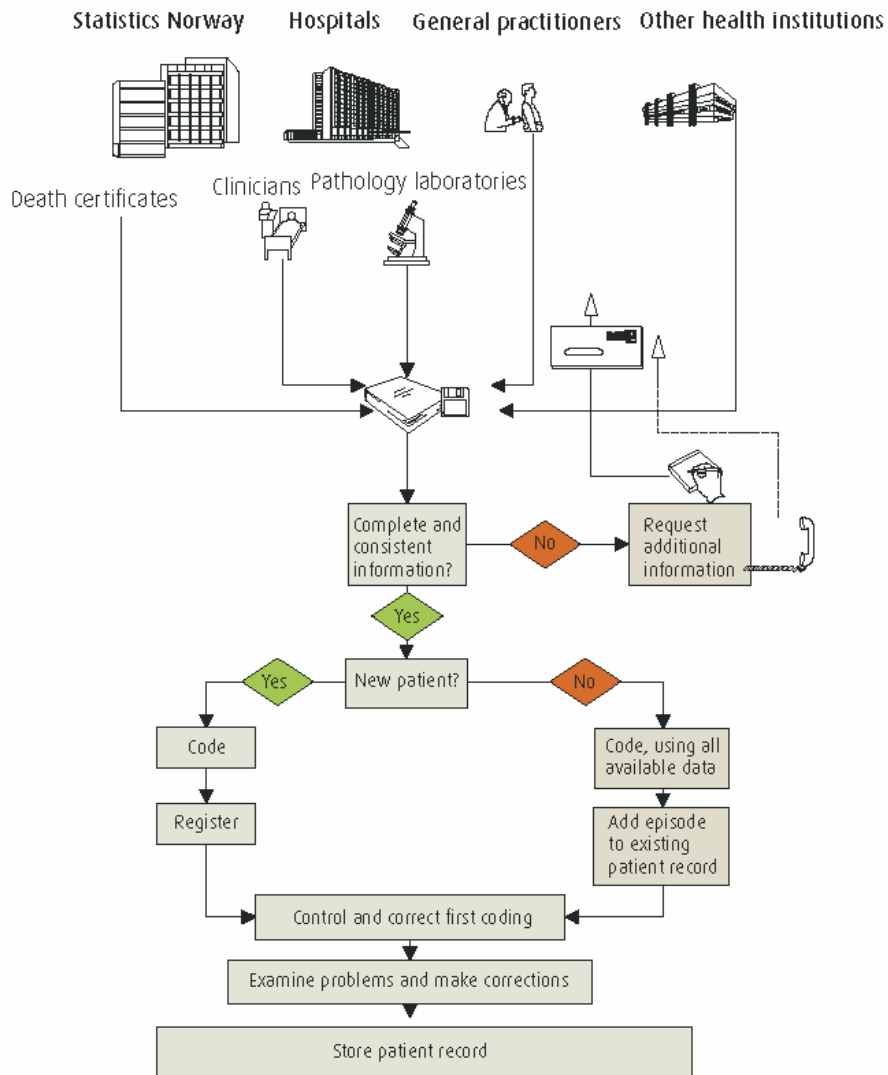
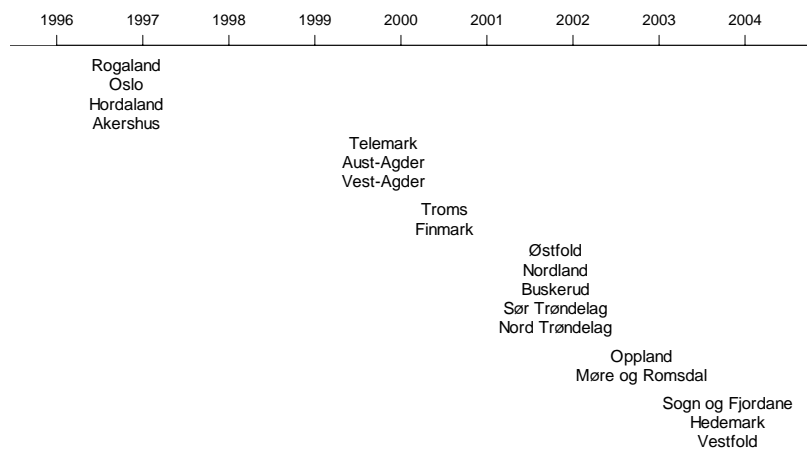


Figure from Cancer in Norway 2005 (54)

Figure 10: Introduction of the Norwegian Breast Cancer Screening Program; year of first invitation round in different Norwegian counties



The program is governmentally funded, supported by a fee from participating women. The actual screening is performed by 26 stationary units and four mobile units (buses). Screening outcome is registered and transferred electronically directly to the local screening databases, and later synchronized with the central screening database at the Cancer Registry of Norway. Post screening examinations of possible cancers are performed at 17 central breast imaging centres at university or county hospitals. The general organisation and logistics of the NBCSP is in accordance with the principles set by WHO (55), with a quality assurance manual and close follow-up from the Cancer Registry of Norway.

3.3 Combining data from different sources

The Norwegian Breast Cancer screening database is designed mainly for the running organization of the Norwegian Breast Cancer Screening Program, but it includes complete and precise information on all invitations, appearances, tests results, possible tumours and results from a questionnaire regarding former screening experience, hormone therapy use etc. The information is stored in separate tables with unique identifications. Data for papers II-IV of this thesis was taken from the screening database and combined with the Norwegian Cancer Registry data,

Papers II-IV in this thesis include screening data from 1995 through 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.

With similar cancer registries, personal identification numbers, mandatory reporting of cancer cases, and governmental screening programs, the Nordic countries (Sweden, Denmark, Finland, Iceland and Norway) are well suited for joint cancer studies. In paper I, we used aggregated screening data from all the Nordic Countries.

with complete registration until 31/12/2002. We combined the individual tables and drew the data directly from the database using SQL commands. Later data were summarized using the S-PLUS statistical package (56), and analyzed in the R statistical package (57). Eligible women received a new invitation to mammography screening 16 to 24 months after their previous screening (with most women receiving their invitation 22-23 months after previous screening). To limit the possible bias from external screening initiated by the screening reminder, all observations

were stopped (censored) two days after the new invitation were mailed (or on death, emigration or after two years of observation for women passing the NBCSP upper age limit of 69 year of age).

For paper I, summarized breast cancer and population figures were collected from all

the Nordic Cancer Registries, and analyzed in the S-PLUS statistical package. In the analysis, data were restricted to the period 1978–1997, except for the Icelandic data, where the years 1973–1977 were added to obtain a sufficient number of cases for the analysis.

3.4 Estimating the effect of breast cancer screening on incidence rates

As for most cancers, breast cancer incidence varies greatly with age. In addition, we have seen distinct changes in breast cancer incidence the last decades, so all analysis of breast cancer incidence rates must take into account both variations in age and calendar time (cohort). This is usually done with Age-Period-Cohort models (58;59), splitting the observations in different age, period and cohort intervals. Having calculated the observed number of cases and person years in each combination, estimates are usually found by maximizing the likelihood of a Poisson regression model. To estimate the effect of screening and deduce cohort estimates corrected for screening activity, we added three variables to the usual Age-Period-Cohort model:

1. As women enter a screening program, screening examinations detecting pre-clinical cancers increase the observed incidence. To estimate this effect in a

population, a specific variable, *screen₁*, modelling the initial effect of a screening program, is used for the two first years a woman is in an official screening program

2. With continued screening, the time of diagnosis is moved to an earlier point in time for some of the observed women. As breast cancer incidence increases with age, a shift in time of the diagnosis increases the observed incidence. To estimate this effect, a specific variable, *screen₂*, modelling the effect of continued screening, is used for each woman in an official screening program after the initial two years
3. As women leave a screening program, the incidence is expected to drop, as some of the expected breast cancer cases already will have been detected at screening. To estimate this effect, a specific variable,

$screen_2$, modelling the effect of continued screening, is used for each woman the first five years after leaving an official screening program

With the added variables, the modelled incidence rate, R_{ap} , for age group a in calendar period p , can be written as:

$$R_{ap} = \exp \left(\begin{array}{l} A_a + D \cdot p + P_p + C_c + S1 \cdot screen_1 \\ + S2 \cdot screen_2 + S3 \cdot screen_3 \end{array} \right)$$

where:

- A_a is the age component for age group a
- D is a common drift parameter (58;59)
- P_p is the non-linear period component of period p
- C_c is the non-linear cohort component of cohort c
- $S1$ is an indicator for the proportion of the population entering a screening program, modelled as the proportion of the population being in the two first years of a screening program

- $S2$ is an indicator for the proportion of the population continuing in a screening program, modelled as the proportion of the population being in a screening program following the two first years
- $S3$ is an indicator for the proportion of the population recently leaving a screening program, modelled as the proportion of the population having left a screening program the past five years.

Applying this new model on breast cancer and screening data, both the effect of screening programs on breast cancer incidence and screening corrected breast cancer incidence trends can be deduced by maximum likelihood estimation.

3.5 Estimating breast cancer progression: the Markov model

Even though nearly all breast cancers found on mammography screening are treated, data from screening trials/programs can shed light on screening test sensitivity [STS] and time in screening detectable phase, so mean sojourn time [MST] (Figure 11), utilizing variations in

breast cancer incidence caused by mammography screening. The basic model of screening related cancer progression is a three state model with women going from a state of “no screening detectable cancer”, through a stage with “pre-clinical cancer visible on

mammograms”, to a stage with “clinical cancer” (Figure 12).

One key assumption of the model is the progression of tumours from “no screening detectable cancer”, through “pre-clinical cancer visible on mammograms” to “clinical cancer”. Some cancers may never progress to clinical disease, while others will never be visible on mammograms, even after clinical detection. Still, moderate levels of both these instances probably constitute only minor problems for the overall model, since the first will be modelled as having very long sojourn time, while the latter will be modelled as having low STS.

Assuming the Markov property (60;61) of conditional independence of prior states, the three stage cancer progression model can be described as a Markov Model. Several authors have applied and extended the literature on Markov models of cancer screening (62-64). Examples include models with additional tumour stages (65), and different distributions for tumour transition times (66).

Under a Markov Model, transition probabilities can be deduced solving a special set of differential equations, so called Kolmogorov’s equations (60;61). Applying this on the standard Markov model of cancer progression, the probability of detecting a cancer at screening can be calculated by multiplying STS with the probability having a cancer in screening detectable phase, given no prior clinical cancer. As explained in Duffy et. al. 1995 (67), this give the following formula for the expected number of cancer cases detected at screening:

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

Where:

- n_s is the number of women screened
- STS is screening test sensitivity
- $\lambda = \frac{1}{MST}$,
where MST = mean sojourn time
- J is the incidence of pre-clinical disease per time unit (typically one month)

Figure 11: Illustration of terms related to screening evaluation:; sojourn time and lead time

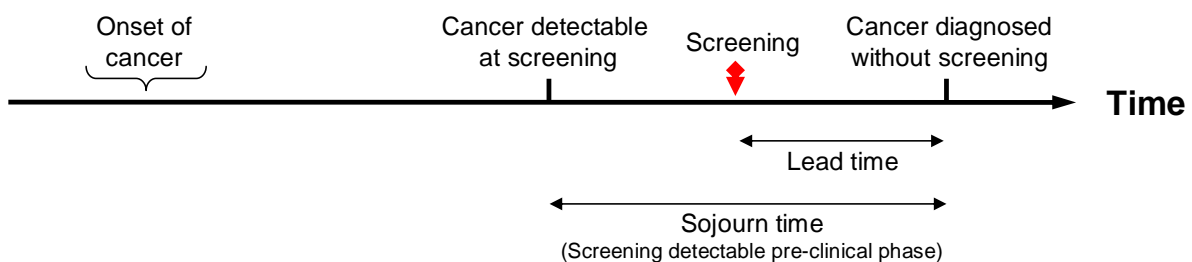
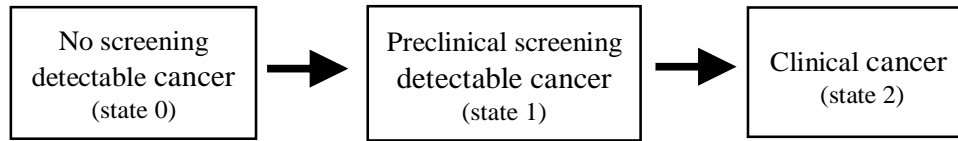


Figure 12: A simple Markov model for breast cancer screening

T is the age at screening

A slightly more intuitive formula is the unconditional approximation:

$$\begin{aligned}
 P(\lambda, STS) & \\
 & \approx n_s \cdot \left(J \cdot \frac{1}{\lambda}\right) \cdot STS \\
 & = n_s \cdot J \cdot MST \cdot STS
 \end{aligned}$$

Which for J and T values relevant for mammography screening give a very good approximation.

The expected number of interval cancer cases in a short time unit, $\langle t_i - 1, t_i \rangle$, 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 (1;67):

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

where c is the number of cancer cases that were detected at screening. In practice, c can be estimated by the formulas above and the

number of persons under observations in the given time interval.

Combining this with screening data on number of cases at screening and during the following interval, non-linear min square regression or maximum likelihood estimates can be deduced (1). In addition to the classical estimation techniques, estimates can also be deduced using the Bayesian theory of “non-informative” Gibbs sampler, but the overall differences compared to maximum likelihood estimates are often small for practical purposes (1).

Another approach is the MISCAN simulation model, used in many papers from a research group in the Nederland’s (63;68). Even though the estimation of parameters is quite different, the practical differences are relatively small as also this model is a stage wise model based on the Markov property.

Even though the basic Markov model of cancer progression is widely used, the models applied are in many ways fairly simple and only partially utilize biological knowledge.

Lately, there has been a trend towards more advanced approaches, including the utilization of many different data sources and more complex models. This typically involves modelling tumour growth using more stages or modelling cancer development as a continuous function of time, using maximum

Chose of estimating method

In the choice between different estimation techniques, several desired properties have to be taken into account. The estimates should be as unbiased as possible, have small variation, and should not be very susceptible to possible erroneous outliers in the dataset.

In the work leading up to paper II, both maximum likelihood [ML] and non-linear least square regression [NLSR] estimations were tested on the NBCSP data. In studying the model fit, we found substantial departures, especially in the start of the interval (paper II; Figure II). In practice, the start of the interval can be influenced by women hesitating to consult their general practitioner shortly after a negative screening examination. Hence, estimates were calculated with and without the data from the first two months following screening. Comparing ML and NLSR estimates with and without the two first months, we found that ML estimates were more influenced by the two first months than the NLSR estimates. Hence, NLSR was chosen in paper II, to limit the impact of the

likelihood estimations or different sorts of simulation techniques. Examples of this include the US National Cancer Institute [NCI] financed CISNET collaboration (69-71), and the new model proposed in paper IV of this thesis.

two possible biased data points at the start of the interval. As for paper III, the differences were minor, and ML estimation chosen.

Working with paper III, a closer investigation of the choice of weights in the non-linear regression were done. With the relatively good model fit in paper III, the practical differences were minimal, but going back to the data used in paper II, an interesting problem surfaced. In many applications, weights are calculated as a function of the expect standard deviation of each data point. With known standard deviations this is a relativity safe choice. However, when estimating the standard deviation, this choice can be problematic as the weights are dependent on the estimated values. In paper II, the weights are repeatedly updated in parallel to the parameter values. As a result, the square difference between the observed and the expected values is not minimized using a fixed weight, but with weights depending on the estimated values. As large

weights produce smaller weighted mean square differences, the estimates are biased towards estimates giving large weights. In most applications, this would probably have a minimal effect on the estimated values (as for the data of paper III), but with moderate model fit the effect on the estimates in paper II was considerable. Hence, future studies should standardize the weights, ensuring that the sum of weights is constant for different parameters.

Even though this is an optimal weight for the correctly specified model, we are not guaranteed the best estimates for situations with a miss specified model. Weights using inverse variances take into account the different random variations of each data point by weighting them according to their random variation. This is a good choice for many applications, but in the Markov model of tumour progression, the observations come from two different sources: data on interval cancers and data on screening cancers. When inverse variance weights are used, the overall

weight of the interval data will vary according to the rather arbitrary choice of intervals for the post-screening period. Using many small time intervals for the post-screening period, estimates would mainly be based on the interval data, while fewer intervals for the post-screening period will give a larger overall weight to the screening data. Applying our new knowledge, weights in the NLSR should make the model robust against outliers and balance the weight given to both the screening and interval part of the data. Defining I_i is an indicator for screening data (e.g. $I_i = 1$ if data point nr i is related to screening data, and 0 otherwise), an example of standardize weights balanced between screening and interval data are:

$$w_i = \frac{1}{(\text{standard deviation}_i)^2} \cdot \left[\frac{1}{2} \cdot I_i \cdot \frac{1}{\sum_{\text{all data points } [i]} I_i \cdot \frac{1}{(\text{standard deviation}_i)^2}} + \frac{1}{2} \cdot (I_i - 1) \cdot \frac{1}{\sum_{\text{all data points } [j]} (I_j - 1) \cdot \frac{1}{(\text{standard deviation}_j)^2}} \right]$$

Quantities estimated from Markov models of cancer progression

From the basic Markov model of mammography screening (Figure 12), two central quantities are estimated; mean sojourn time [MST] and screening test sensitivity [STS].

While lead time describes how far screening moves the diagnosis, sojourn time is a measurement of the time a tumour spends in a pre-clinical screening detectable phase. Mean sojourn time and the distribution of sojourn times are functions of tumour progression and the given screening test, while lead time also

is a function of screening frequency. Hence, the more general term of sojourn time is usually reported.

As cancer is assumed to be a progressive disease, starting with genetic changes in one or a limited number of cells, screening test sensitivity is not as easily defined (44). Evaluating screening programs, a common definition of STS is the proportion of cancers found at screening divided by the total number of cancers at screening and the those detected during following year (44). This

3.6 Estimating breast cancer growth

For many applications, the Markov models of cancer progression is a substitute for more precise information relating tumour growth and STS at different tumour sizes. There are some observational studies of tumour growth on tumours initially overlooked at earlier mammograms (27-29), or based on tumours in patients refusing treatment (27;28), but these studies are small, and probably influenced by considerable length time bias, as slow-growing tumours spend relatively longer time in pre-clinical stages that are visible on mammograms. Hence, better estimates of tumour growth and screening test sensitivity more directly related to tumour size would be useful. Compared to the studies of overlooked cancers, population based studies greatly increases the number of

definition is, however, questionable, as many cancers found on screening probably would have used several years to become clinically detected cancers without screening. In the Markov model, STS is modelled as a step wise function, going from zero in the first stage of “no screening detectable cancer”, to a level given by the STS variable at the second stage of the “pre-clinical cancer detectable on mammograms”. This definition can in practice be seen as the “number of pre-clinical cancers detectable on mammograms found during one screening examination”.

observed cases and apply data that are probably less vulnerable to potential biases.

Combining the NBCSP standardized tumour measurements with the variations in cancer incidence used in the Markov models, there is considerable information about the growth rates of pre-clinical breast cancers. Spratt et al. 1993 (72) used a variant of a general logistic growth curve with log-normal distributed growth rates on a clinical dataset mostly consisting of overlooked tumours. Combining Spratt’s model with a two-parameter screening sensitivity curve modelled as a logistic function of tumour size, we have developed a new estimation method for tumour growth, which is presented in paper IV. The calculation of the expected

number of cases, for a given parameter set, at screening is done using “back calculations” from the expected number of future cancers. Having deduced the expected number of cases

in different size groups at screening and time intervals after screening, maximum likelihood estimates can be calculated using multinomial and Poisson distributions.

3.7 Estimation in practice: challenges and solutions

When calculating maximum likelihood parameters for the new estimation technique proposed in paper IV, the expected number of cases is deduced by applying probability and tumour size “back calculations” on a large number of future time intervals for each tumour size group. Hence, a huge number of calculations are involved even for finding the expected number of cases for just one parameter set. As a result, finding maximum likelihood estimates through the optimization of the four dimensional log likelihood consumes a considerable amount of computer time, even before finding bootstrap confidence intervals. To ease the practical problems regarding required computing time, several measurements were taken. First, we tried using the C language as an alternative to the higher level R language (57), but with little success, as further investigations revealed that it was the large number of log-normal distribution probability calculations that was the main problem.

To limit the number of probability calculations, a special procedure was used in the final calculations: in stead of calculating each time interval separately, calculations were started with a given tumour size group at screening, and the upper and lower growth rate intervals for the relevant combinations of screening and clinical tumour sizes were calculated. Using this approach, the consumed computer time was reduced considerably, but we still had to setup a dedicated computer to perform the final calculations using several weeks of computing time.

4 Summary of results

4.1 Summary of paper I

*The influence of mammographic screening
on national trends in breast cancer incidence*

Bjørn Møller, Harald Fekjær, Timo Hakulinen, Laufey Tryggvadóttir,

Hans H. Storm, Mats Talbäck and Tor Haldorsen

(European Journal of Cancer Prevention; Volume 14, no 2: 117-128)

As mammographic screening programs aim to reduce the mortality through earlier diagnosis and treatment, successful mammographic screening programs affect breast cancer incidence rates in a population. In practice, there are different effects at the start of a screening program, during a screening program, and after the end of a screening program. The number of future breast cancer cases is important in the planning of future cancer diagnostic and treatment services, and screening is one factor that should be accounted for in predicting the number of future breast cancer cases (8;52). Hence, estimates of the effects of a screening program on observed breast cancer incidence are needed. To quantify the potential effects of mammographic screening programs, a special age-period-cohort (58;59) model with separate variables for the effect of a given proportion of the observed population entering a screening program, continuing screening and having left a screening program was utilized. The model was applied to data from the five Nordic countries: Finland, Denmark, Iceland, Norway and Sweden. In addition the model allowed us to estimate screening corrected incidence trends.

Having the largest population and the longest running screening programs, the best estimates were obtained in Sweden. Swedish breast cancer rates more than doubled (relative risk = 2.20, 95 % CI; 1.8–2.6) for populations first offered screening compared with pre-screening breast cancer incidence. The risk remained elevated (relative risk = 1.34, 95 % CI; 1.2–1.6) with a continued screening program, while the rates dropped (relative risk = 0.68, 95 % CI; 0.6–0.8) when the women left the program. This indicates that screening advances the time of diagnosis, which is a prerequisite to a subsequent reduction in mortality. The effects were considerable, requiring a correction for screening activity when calculating future cancer burdens. Analysis of secular trends, corrected for the influence of screening, showed that the rates in Finland increased by 13.1 % per 5-year period, Denmark 3.1 %, Iceland 2.1 %, Norway 3.7 % and Sweden 1.1 %. There were strong cohort effects in all Nordic countries, and the risk seemed to level off for the youngest cohorts in most of the countries.

4.2 Summary of paper II

Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening; New results

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

(Journal of Medical Screening, Volume 12, no 4: 172-178)

Average time in pre-clinical screening detectable phase, so called mean sojourn time [MST], and screening test sensitivity [STS], are central parameters in the planning and evaluation of breast cancer screening (42;73). New screening techniques, increased use of hormone replacement therapy or the transition from breast cancer screening trials to large scale screening programs may influence both MST and STS. Hence, a three step Markov chain model was applied to data from Norwegian Breast Cancer Screening Program [NBCSP]. With possible problems of opportunistic screening between ordinary breast cancer screening rounds, a special sensitivity analysis of this potential problem was performed.

MST was estimated to 6.1 (95 % CI: 5.1-7.0) years for women aged 50-59 years, and 7.0 (95 % CI: 6.0-7.9) years for those aged 60-69 years. Correspondingly, STS was estimated to 58 % (95 % CI: 52-64 %) and 73 % (67-78 %), respectively. Simulations revealed that opportunistic screening may give a moderate estimation bias towards higher MST and lower STS. Assuming a probable 21 % higher background incidence, due to increased hormone replacement therapy use, MST estimates decreased to 3.9 years and 5.0 years for the two age groups, and STS increased to 75 % and 85 %. Compared to previous reported MST and STS from other screening programs or trials (42;74), these new estimates indicate that the screening detectable phase is longer in the NBCSP, but also that the sensitivity of the screening test is lower. Overall, the NBCSP detects more cancer cases than most previous trials and programs.

4.3 Summary of paper III

Estimating 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, Steinar Tretli

(submitted)

Average time in pre-clinical screening detectable phase, so called mean sojourn time, and screening test sensitivity, are central parameters in both the evaluation and planning of screening programs. Both quantities are usually estimated by a Markov model using incidence data from the first screening round and the interval between screening examinations. Several screening programs do not, however, have full registration of cancers emerging after screening, and the increased use of opportunistic screening over time may raise questions regarding the quality of interval cancer registration.

To estimate mean sojourn time and screening test sensitivity without the use of interval cancers, we used questionnaire data on time since previous screening combined with screening outcome, and deduced required mathematical formulas under the same Markov model as used in paper II. The new approach was applied on questionnaire data for 336 533 women in the Norwegian Breast Cancer Screening Program [NBCSP], using non-linear weighted mean square regression to estimate mean sojourn time [MST] and screening test sensitivity [STS]. In contrast to the method used in paper II, the new method gave a satisfactory model fit.

Assuming a probable 21 % higher background incidence due to increased hormone replacement therapy use, 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 were estimated to 55 and 60 percent, respectively. Hence, estimates from paper II indicating a long mean sojourn time and low screen test sensitivity in the NBCSP were confirmed with a new approach less vulnerable to unregistered mammography screening activity.

Regarding the different approaches to estimate MST and STS, we found that questionnaire data on time since previous screening can be used to estimate mean sojourn time and screening test sensitivity, but that the model is sensitive to relaxing the assumptions regarding the expected breast cancer incidence without screening and constant STS over time.

4.4 Summary of paper IV

Breast cancer tumor growth estimated through mammography screening data

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

(submitted)

Knowledge of tumour growth is important in the planning and evaluation of screening programs, clinical trials and epidemiological studies. Studies of tumour growth rates in humans are usually based on small and selected samples, as almost all diagnosed cancers are treated. In paper IV, we develop a new likelihood based estimating procedure of tumour growth that can be used on population based screening data. The new procedure is based on a tumour growth model where both tumour growth and screen test sensitivity [STS] is modelled as continuously increasing functions of tumour size. The method was applied on cancer incidence data including tumour measurements from 395 188 women aged 50-69 years participating in the Norwegian Breast Cancer Screening Program.

Tumour growth varied considerably between subjects, with 5 % of tumours using less than 1.2 months to grow from 10 to 20 mm in diameter, and another 5 % using more than 6.3 years. The mean time a tumour needed to grow from 10 to 20 mm in diameter was estimated to 1.7 years, increasing with age. STS was estimated to increase sharply with tumour size, going from 26 % at 5 mm to 91 % at 10 mm. Compared to the previously used Markov models for tumour progression used in paper II, the applied model gave considerably higher model fit (85 % increased predictive power) and resulted in estimates directly linked to tumour size.

The study shows that screening data with standardized tumour measurements can give population based estimates of tumour growth and STS directly linked to tumour size.

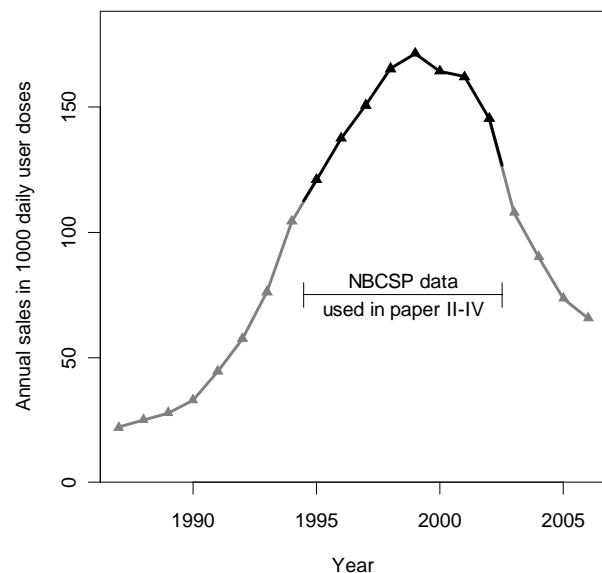
5 Discussion: methodological choices and potential biases

5.1 Bias caused by variations in HRT use

In the absence of screening or new diagnostic tools, the relative risk of cancer for different age groups usually changes gradually over time. Hence, we have modelled breast cancer incidence in paper I by comparing trends in incidence rates for different combinations of cohort and age at diagnosis, correcting for differences in screening activity. The estimated Norwegian trend in breast cancer incidence by calendar time was then applied in papers II-IV to calculate the expected breast cancer incidence without screening. In most instances, this would be a fairly robust estimate, but there is one factor probably influencing recent breast cancer incidence trends of middle aged Norwegian women substantially, namely hormone replacement therapy [HRT]. HRT is a medical drug used for various physical hassles in connection with women's menopause. With an increasing belief in a protective effect of HRT against severe health problems such as cardiovascular morbidity (75), HRT use increased rapidly until the Women's Health Initiative study [WHI] reported an increased risk of both breast cancer and cardiovascular disease in a large randomized trial (76).

In Norway, the sharp increase in HRT use coincided with the introduction of the

Figure 13: Systemic hormone therapy [HRT] use in Norway 1987-2006, in daily doses



Source: Dealer based drug consumption statistics, Norwegian Institute of Public Health

public mammography screening program (Figure 13). The NBCSP was introduced gradually from 1995 to 2004 for women 50-69 years of age (17;77;78), while HRT use increased sharply in the 1990s with a peak in 1999 (79), mainly affecting women during the first years following menopause (77). As a result, the screening variables in paper I are likely to pick up some of the increased breast cancer incidence due to HRT use, giving some bias in the estimated values.

When writing paper I, we did not fully realize the possible impact of increased HRT use.

Rossouw et al. had in 2002 published the report from Women's Health Initiative study, estimating that current HRT use increases the risk of breast cancer by 26 % (76), but we thought that an effect of this magnitude was of less importance for our estimates. In 2004, Bakken et al. estimated the relative Norwegian risk for current versus never users of HRT to 2.1, based on the Norwegian Women and Cancer [NOWAC] study (17). In addition, Bakken et al. pointed out the large increase in Norwegian HRT sales during the 1990s, with the largest increase occurring around the start of the NBSCP. Overall, they estimated the HRT attributable proportion of breast cancer cases to 27 % for Norwegian women 45-64 years of age (17). Hence, our estimated values could be essentially biased by not taking into account the increased HRT use in the Norwegian population.

Looking back at the approach used in paper 1, the *screen₁* variable, modelling the effect of entering a screening program, was in Norway probably a good proxy for increased HRT use, coinciding in regards of both calendar time and age groups. While the NBCSP had only been established in some counties during the analysed time frame, HRT use was probably widely distributed. Hence, HRT use was probably even more widespread than screening, and occurring in the same age-period-cohort combinations. With no variable for increased HRT use, HRT use is likely to

account for a considerable proportion of the *screen₁* estimate, explaining the higher estimated effect on breast cancer risk of entering the Norwegian screening program than the screening programs introduced in the other Nordic countries.

In the other Nordic countries, the effect on the *screen₁* variable is probably smaller, as screening was introduced earlier. On the other side, the increase of HRT use probably coincides more with the continued screening variable, *screen₂*, giving biased *screen₂* estimates. An indication of this is found in Jonsson et al. 2005 (80), analysing the increased incidence with time since screening. Jonsson et al. find that there was a permanently increased incidence among continuously screened Swedish women, that do not fit well with existing lead time estimates. On this basis, Jonsson et al. concludes that "The reason appears to be a high presence of small sub clinical breast cancers in the female population, especially at ages 50-59 years. Many of these tumours would never have been disclosed without mammography". It may be true that there are substantial problems with over-diagnosis, but the figures also revealed one interesting fact: the breast cancer incidence for women 50-59 and 60-69 years of age increased with initial screening, fell to a modestly increased level after the initial screening, and then seemed to increase again with time since first screening.

This is surprising, since an increasing incidence of continuously screened woman with time since first screening neither fit with the assumption that repeated screening examinations reduce the pool of screening detectable cancers, nor the expected effects of over-diagnosis. As there is little evidence of a considerable increase of breast cancer risk after mammography related radiation (44), a likely explanation is the increased HRT use by time.

Working with paper II, we were more attentive to the possible impact of increased HRT use. To correct for this potential bias, we also calculated our estimates adding 21 % to the assumed background incidence on the basis of information regarding increased breast cancer risk and HRT sale figures found in the aforementioned study by Bakken et al (17). Progressing further with paper III and IV, our belief of an essential HRT effect convinced us that the correction for increased HRT use would probably give more correct estimates.

Still, there exists a great uncertainty regarding the influence of HRT on Norwegian breast cancer rates. The differences in results from the WHI and NOWAC studies are not fully understood. Some of the differences may be attributable to different pharmaceutical

products used (81), or perhaps reflect a possible interaction with earlier oral contraceptive use as indicated by Lund et al (81).

Recently, Hofvind et al. (82) estimated the population attributable risk of HRT in the NBCSP during 1996-2004 to 20 %, while Ravdin et al. (83) observed a marked drop in the US breast cancer rates after the reduction of HRT use following the WHI trial. This strengthens the theory of a considerable increased Norwegian breast cancer risk due to increased HRT use, but further research is still needed to quantify the effects of increased HRT use.

Bakken et al. (17) studied HRT use in different age groups, and reported a peak among Norwegian women aged 50-54 and 55-59 years, with less use HRT use in the 45-49 and 60-64 age groups. As a result, we may have over-corrected the background incidence in paper II-IV for the 60-69 age group, while the opposite may be the case for the 50-59 age group. Overall, the new model in paper IV is more robust than the traditional Markov model used in papers II and III. This is probably a consequence of the model in paper IV including data from more sources, hence limiting the bias by a miss-specified background incidence.

5.2 Bias from unregistered screening

With estimating techniques exploiting the variations in cancer incidence related to screening, it is important to separate interval cancers appearing at private opportunistic screening from interval cancers detected clinically. As the Cancer Registry of Norway includes reports on the background of each interval cancer, we have investigated the possibility of identifying cancers detected at opportunistic screening. The indication for mammography was, however, often insufficiently, and sometimes even incorrectly, reported. Hence, the background for many interval cancers is unknown, and the applied interval cancer rates in papers II and IV can be contaminated by opportunistic screening.

An ideal solution would of course be mandatory registration of all mammography screening, including examinations performed at both public and private institutes. Work towards this is underway in Norway. On the other hand, many countries worldwide will probably never regulate the reporting of mammography screening performed at private

institutions. This was a minor problem in the start of the international screening era, but has become increasingly problematic as mammography has become a commonly available health service. In order to thoroughly discuss the possible impact of unregistered screening, there is a need for a quantification of the problem regarding unregistered opportunistic screening.

Inquiries from the Cancer Registry of Norway have indicated that approximately 10 % of women in the NBCSP target group are screened each year. These estimates are, however, uncertain, since we have limited knowledge of the age groups involved and the distribution between NBCSP attendees and non-attendees. Hence, a more direct way of estimating the level of opportunistic screening would be beneficial. In the appendix (page 57), a new approach for estimating the likely volume of unregistered screening, based on the observed DCIS frequency, is suggested.

Correcting for unregistered mammography screening

Applying the new estimation approach given in the appendix (page 57), estimates from paper II and IV can be adjusted for an estimated level of opportunistic screening. Adjusting the observed incidence of interval cancers, by removing the assumed incidence caused by opportunistic screening (Figure 14), we find that both the Markov model estimates given in paper II, and the new estimation method used in paper IV, are vulnerable to estimation bias caused by unregistered opportunistic screening (Table 1 and Table 2). Hence, the new robust approach using time since previous screening presented in paper III is a good alternative. The estimated values are, however, close to the estimates in paper II, with even longer MST and lower STS. This was surprising, and

strengthens the estimates given in paper II. As a consequence bias from opportunistic screening is probably smaller than estimated in the appendix, or estimates in paper III are biased by different mammography quality between the NBCSP and the earlier mammography examinations.

As for the new model presented in paper IV, we find that mean tumour growth and STS are fairly robust for opportunistic screening, but that the variation in growth rates is estimated to zero when correction for opportunistic screening (Table 2). As there is sufficient evidence for an individual variation in breast cancer tumour growth rates (24;27;29), this is probably an effect of an over-correction.

Figure 14: Breast cancer interval rates with and without correction for the estimated opportunistic screening activity. Monthly observed rates are plotted as “o”, while the corrected estimates are marked with “+”. Smoothed vales are given with “—” and “- -”, respectively

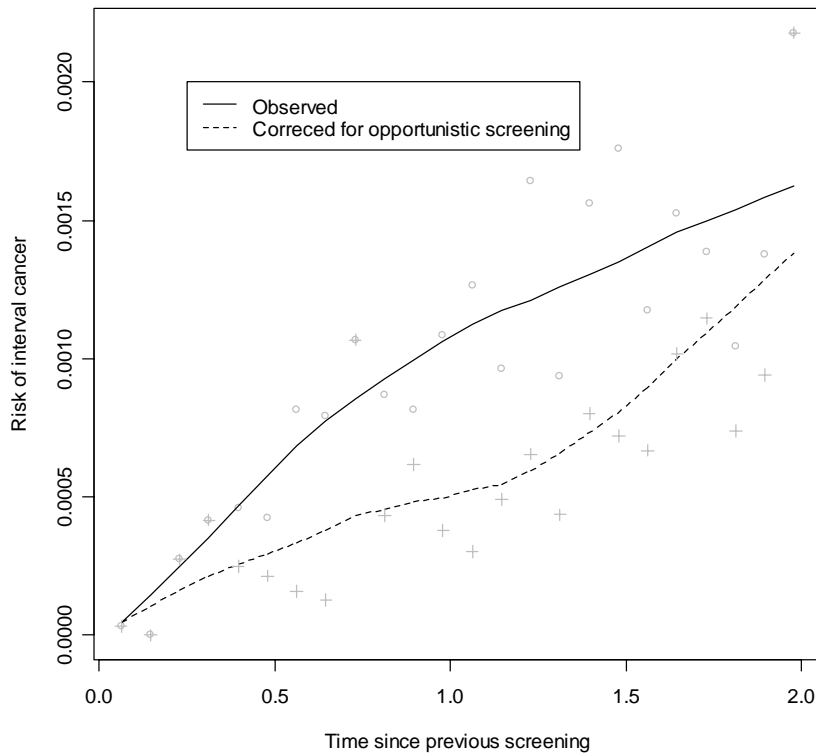


Table 1: Estimates from paper II with and without correction for a possible bias by opportunistic screening, all age groups combined. All estimates are corrected for an assumed 21 % increase in background incidence attributed to increased HRT use

	MST (years)	STS (%)
Without correction	4.2	82
With correction	3.5	99

Table 2: Estimates from paper IV with and without correction for a possible bias by opportunistic screening, all age groups combined. All estimates are corrected for an assumed 21 % increase in background incidence attributed to increased HRT use

Model	<i>Time (years) used from 10 to 20 mm:</i>		Screening test sensitivity at:		Indicators of potential screening efficacy:	
	Mean	Standard deviation	5 mm	10 mm	Mean sojourn time (years)	Proportion of tumours visible on screening
Original data	1.7	2.2	26	91	2.9	0.95
Correction for opportunistic screening	1.7	0.0*	22	84	2.4	0.94

* Variance parameter (α_2) estimated to zero

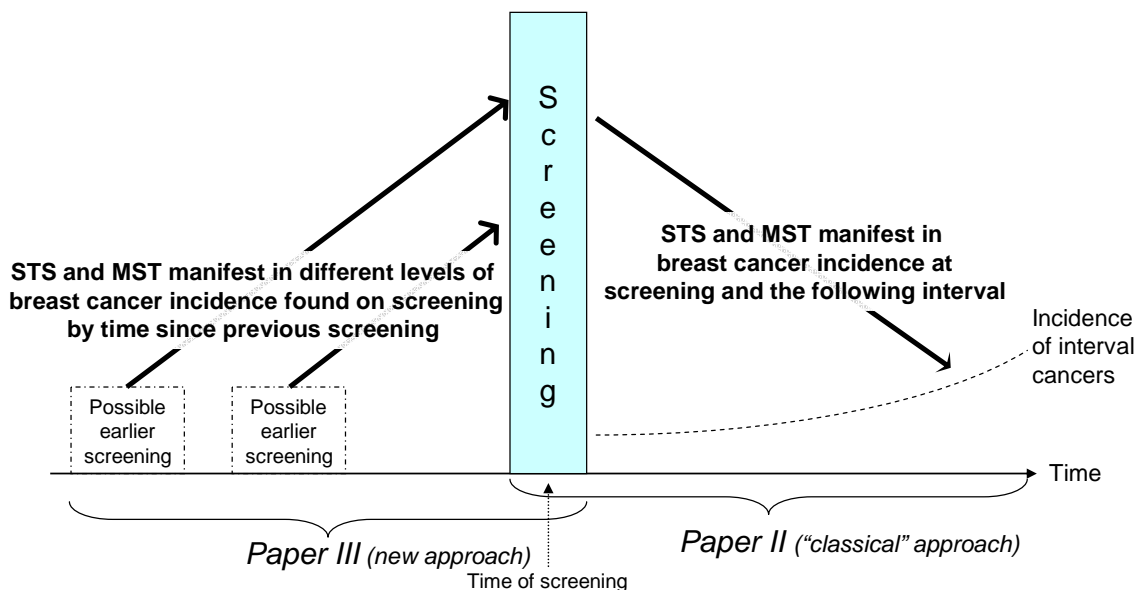
5.3 MST and STS in the NBCSP

In this thesis, MST and STS are estimated based on two different datasets. In paper II, MST and STS are estimated applying data on screening and interval cases, while a new approach in paper III uses time since last screening examination combined with the screening result (Figure 15). In practice, estimated MST and STS vary considerably when different approaches are used. In addition, there are several possible choices for central assumptions as the expected background incidence without screening and possible exclusion of DCIS cases. For women 50-59 years of age, MST estimates varies from 3.9 to 7.2 years, and STS estimates from 52 to 75 %. Correspondingly, MST estimates varies from 5.0 to 8.6 years, and STS estimates from 58 to 85 %, for women 60-69

years of age (paper II and III). Hence, there is substantial variation in the estimated values, but the estimated expected incidence on first screening is fairly stable. This is probably an effect of the estimated MST and STS being highly negatively correlated (Figure 16), making it difficult to separate the possible effects of either a high MST or a high STS.

The large variation in estimated values makes it difficult chose the best estimates for the NBCSP MST and STS. As the approaches in papers II and III are based on the same assumptions, estimates can also be derived combining all the available data. Applying both interval data and data on time since previous screening, MST is estimated to 5.4 years for women 59-59 years of age, and 6.1

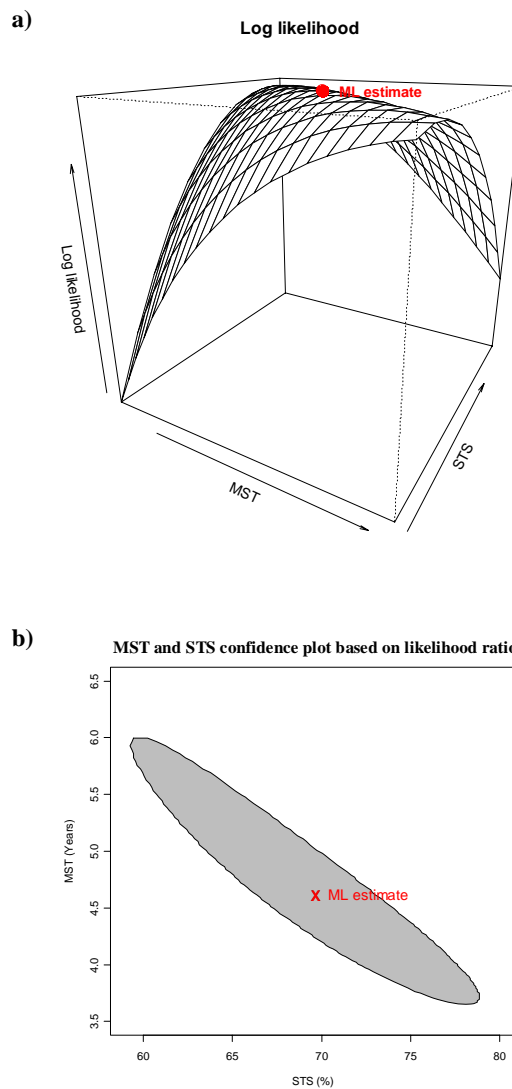
Figure 15: Illustration of how mean sojourn time [MST] and screening test sensitivity [STS] affect breast cancer incidence at different stages of a screening program. As indicated at the bottom of the figure, papers II and III utilize different parts of the observed incidence variations to estimate MST and STS



years for women 60-69 years of age, when the assumed background incidence is increased with 21 % because of increased HRT use (Table 3). Correspondingly, STS is estimated to 55 % and 68 %. Comparing the combined estimate, with the original estimates seen in papers II and III, it may seem strange that confidence intervals are not always narrower than the original confidence intervals, even when more data is used. Studying model fit, we find that the combined model has slightly less non-optimal fit, as the overall estimate is a compromise between the two original sets of estimates. Hence, the decreasing standard error due to more observations is compensated by a less optimal model fit. Even when no overall reduction in standard errors is achieved, the pooled estimate is probably our best approximation of the NBCSP MST and STS.

To further investigate the NBCSP STS, we have also applied a third approach using the observed drop in cancer cases from the first to the second screening round. This approach is based on the assumption that breast cancer tumours probably seldom or never regress. Under this assumption, all cancer cases found on first screening would either surface before the second screening round, or still be in pre-clinical screening detectable phase by the time of the second screening round. Hence, the relative pool of screening detectable breast cancer at second screening could be estimated

Figure 16: Illustrations of the strong correlation between MST and STS estimates derived from the “classical” Markov model of mammography screening: a) The log likelihood with a distinct ridge along a combination of increasing MST and decreasing STS b) Confidence region for MST and STS parameters, with a distinct oval form



Data: NBCSP women 50-59 years of age, not corrected for increased HRT use

by the number of cancers detected at the first screening, combined with STS and the drop in new cancer cases between the two screening examinations (relative to the expected incidence without screening). Exploiting this,

Table 3: MST and STS estimated using different parts of the NBCSP data, when the assumed background incidence is increased with 21 % because of increased HRT use

	MST estimate (years)		STS estimate (%)	
	50-59 years	60-69 years	50-59 years	60-69 years
Using interval data (paper II) *	3.9 [3.2 – 4.2]**	5.0 [4.3–5.5]**	75 [70 – 82]**	85 [80 – 90]**
Using time since screening (paper III)	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

formulas estimating STS can be easily deduced, and STS estimated by:

$$STS = \frac{PCS_1 - PCS_2}{PCS_1 - \int_{\text{Time of first screening}}^{\text{Time of second screening}} [h_b(t) - h_e] dt}$$

Where PCS_1 is the proportion of women with a cancer detected on the first screening examination, PCS_2 is the proportion with a cancer detected on the second screening examination, h_e the expected risk of cancer without screening, and $h_o(t)$ the observed risk of interval cases at time t .

In practice, a large drop from the first to the second screening round, combined with few

interval cancers, would give a high STS estimate, as it indicates a very effective first screening round (and vice versa). This approach can be seen as a validation of our earlier STS estimates, and the new estimates confirms the relatively low STS estimated in papers II and III (Table 4).

The new approach, based on the observed drop in cancer cases from the first to the second screening round, has, however, some weaknesses as a general estimation method. STS is likely to increase with new screening rounds, as previous mammograms become available and staff gets more training. As a result, the estimates are biased towards lower STS estimates.

Table 4: A validation of STS estimates using data from repeated screening examinations

	Age group	
	50-59 years	60-69 years
Original data	49 %	71 %
With 21 % added to the assumed background incidence attributed to increased HRT use.	69 %	79 %

Comparison with earlier studies

Tabar et al. (42) estimated MST to 1.7, 3.3, 3.8 and 2.6 years for women who were 40-49, 50-59, 60-69 and 70-74 years of age, respectively, in the Swedish Two County trial. The corresponding STS were estimated to 86 %, 92 %, 94 % and 100 %, correspondingly.

Similarly, Paci et al. (74) estimated MST and STS from both the US “Health Insurance Plan” (38), the Italian “Florence District Program” (74), and the Swedish “Two county Study” (42), based on the same Markov model assumptions. MST estimates ranged from 1.2 to 6.5 years, increasing with age, while STS ranged from 69 % to 95 %, also increasing with age (74).

While most studies use regular estimation techniques for Markov methods, a Dutch research group used the MISCAN simulation model (63;84), where the data is fitted through systematic testing of different parameter values in a simulation model.

Applying the model on screening data from Utrecht and Nijmegen screening projects, MST was estimated to range from 1.6 to 5 years, increasing with age. Correspondingly, STS was estimated to 70 % for DCIS and invasive cancers under 10 mm, and 95 % for invasive cancers above 10 mm.

In summary, MST for mammography screening of women 50-69 years of age has typically been estimated to around four years, while STS has been estimated to just below 90 %. Compared with earlier studies, the new NBCSP estimates of MST are considerably higher, while the STS estimates are lower, even when chosen the most conservative estimates. Independently of our assumptions, the new estimates confirm the increasing MST and STS with age as estimated in earlier screening trials/programs.

The increased MST implies that patients spend more time in the pre-clinical screening detectable phase. Studying data from the

Norwegian Haukeland database with tumour measurements of clinical cancers taken before the NBCSP (used in paper IV), we find that mean clinical tumour size has fallen over time, and is less than reported from the Swedish trials (42). Hence, the increased time in screening detectable phase (MST) is probably an effect of the screening test diagnosing smaller tumours, rather than later clinical detection. During the last years, efforts have been made to increase the quality of mammographic images as well as other aspects of mammography screening (85;86). These improvements are likely to decrease the minimum size of tumours visible on mammograms, 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 cases (87). Comparing NBCSP with the Nijmegen and Utrecht screening projects (63), the proportion of DCIS in the first screening round was 17 % and 13 %, respectively. Similarly, the proportion of invasive cancer cases 20 mm or larger in the first round was 19 % for the NBCSP, compared to 28 % combined for the Nijmegen and Utrecht screening projects.

Comparing early screening performance indicators with results included in the Lyng et al. (88) review, the NBCSP has a generally

high detection rate and proportion of DCIS cases, but also very high levels of interval cancers. This corresponds well with a unstable screening test capable of detecting small pre-clinical tumours.

The low sensitivity may seem strange in the light of the ongoing efforts to increase mammographic quality. An explanation for the low STS could be widespread HRT use, as HRT is known to reduce STS (89-91). In Norway, the use of HRT was common during the study period (17), especially among women 50-59 years of age (77). In addition, one should note that the STS in the Markov models is not an absolute quality measurement, and that the expected number of diagnosed cancer cases at each screening round was higher than reported in most previous studies. Hence, improved routines may enable the detection of additional small tumours, but simultaneously yield lower STS as more tumour are considered to be in pre-clinical screening detectable phase.

To summarize, the NBCSP detects more cancer cases than the early randomized trials, even low STS estimates indicates that many cancers are overlooked at each screening. Longer MST gives good hope of increased survival through earlier diagnosis, but can also increase the level of over-diagnostics, as more women are diagnosed that without screening may never had developed clinical

breast cancer in their remaining life time. The high interval rate and low STS estimates,

indicate substantial opportunities for increasing screening quality.

5.4 Comparing the Markov model and the new growth model (of paper IV)

Comparing the Markov model of papers II and III with the growth model of paper IV, we find that STS have been defined in two very different ways. STS in the Markov model is a measurement of screening test reproducibility, defined as “the number of pre-clinical cancers visible on mammograms found on one screening examination”, while the STS in the growth model of paper IV is a direct test sensitivity measurement for tumours of a given tumour diameter (“the proportion of tumours visible on screening at X mm”). In addition, paper IV estimates the “proportion of tumours visible on screening” defined as “the proportion of tumours that at one point is visible on a mammogram before clinical detection”. Hence, the definitions differ, with few possibilities for direct comparisons.

On the other hand, MST has a more similar definition in all papers. Comparing MST estimates of paper II, III and IV, we find that MST estimates is considerable lower in paper IV. With data from the same screening program, this could seem like a contradiction. Fitting the “classical” Markov model of paper II, to simulated data based on the model and estimated parameters found in paper IV, MST still differs considerable from the estimated

values of paper IV. This is probably an effect of the different STS and model definitions, relating the sojourn time to different populations. In the growth model of paper IV, each tumour’s sojourn time begins at a given individual tumour size, from where the tumour is visible on screening with 100 % sensitivity until clinical detection. On the other side, sojourn time in the Markov model is followed by a given STS. Hence, sojourn time, and MST, have different interpretations in the two models, resulting in different estimated values. Look at simulated screening scenarios, results are much closer with better correspondence.

Applying the two different models, the tumour growth model of paper IV estimates biological properties that are not available from the Markov Model, but could also serve as an alternative to the Markov Model for simulating the outcome of key questions related to different screening designs. With estimates directly connected to tumour size, and easy separation of MST and STS, the growth model of paper IV is probably a good alternative to the Markov Model for many applications when precise tumour measurements are available. On the other

hand, the Markov Model is a better tested model, with less data requirements, yielding results more directly comparable with earlier studies.

6 Conclusions

- Correcting for different screening activity, the risk of breast cancer has increased considerably in all the Nordic countries from 1978–1997, with a possible stabilisation of risk for the youngest cohorts. The largest increase has been observed in Finland (13 % per 5-year period), and the smallest in Sweden (1.1 % per 5-year period). In Norway, breast cancer rates increased with a mean of 3.7 % per 5-year period.
- It is possible to estimate the effect of service screening programs on breast cancer incidence, based on cohort data and knowledge of screening patterns. Estimates from Sweden indicate that the Swedish screening programs has had a large effect on breast cancer incidence, with the risk more than doubling at the introduction of the screening programs (relative risk = 2.2), followed by an increase risk of 35 % with continued screening, and a 32 % reduced risk the first five years following screening. The estimates are, however, uncertain due to the lack of a correction for a likely swift change in risk due to in hormone therapy use.
- Breast cancer diagnosis is probably moved further (larger mean sojourn time) in the NBCSP than in most earlier screening programs and trials, but probably with a relatively low screening test sensitivity.
- Collected data on time since previous screening examination can be used to estimate mean sojourn time and screen test sensitivity, based on data from only one screening examination without information concerning interval cancers (paper III).
- Tumour growth and screening test sensitivity, directly linked to tumour size, can be estimated using data from large population based screening programs, combining modern computer power and new estimation approaches (suggested in paper IV).
- The mean time a tumour needs to grow from 10 to 20 mm in diameter was estimated to 1.7 years, increasing with age. Tumour growth was estimated to vary greatly between subjects, while screening test sensitivity is estimated to increase sharply with tumour size, increasing from 26 % at 5 mm to 91 % at 10 mm.

7 References

- (1) Prevost TC, Launoy G, Duffy SW, Chen HH. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *Am J Epidemiol* 1998; 148(6):609-619.
- (2) Day NE. Overdiagnosis and breast cancer screening. *Breast Cancer Res* 2005; 7(5):228-229.
- (3) Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005; 55(2):74-108.
- (4) Stewart BW, Kleihues P. World cancer report. Lyon: IARC Press, 2003.
- (5) GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide. IARC CancerBase No. 5. Lyon, IARC Press, 2001, 2001.
- (6) Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H. Trends in cancer incidence and mortality. Lyon: IARC press, 1993.
- (7) Ursin G, Bernstein L, Pike MC. Breast cancer. *Cancer Surv* 1994; 19-20:241-264.
- (8) Moller B, Fekjaer H, Hakulinen T, Tryggvadottir L, Storm HH, Talback M et al. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 2002; 11 Suppl 1:S1-96.
- (9) Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AM, West DW et al. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993; 85(22):1819-1827.
- (10) Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973-1997. *Int J Epidemiol* 2005; 34(2):405-412.
- (11) Leung GM, Thach TQ, Lam TH, Hedley AJ, Foo W, Fielding R et al. Trends in breast cancer incidence in Hong Kong between 1973 and 1999: an age-period-cohort analysis. *Br J Cancer* 2002; 87(9):982-988.
- (12) Cancer Registry of Norway. Cancer in Norway 2004. <http://www.kreftregisteret.no> . 2. 11-9-2006.
Ref Type: Electronic Citation
- (13) Wang H, Thoresen SØ, Tretli S. Breast cancer in Norway 1970-1993: a population-based study on incidence, mortality and survival. *Br J Cancer* 1988; 9(77):1519-1524.
- (14) Thoresen SØ. Prøveprosjekt med mammografiscreening i fire fylker; organisering og gjennomføring. *Norsk Epidemiologi* 1997; 7(2):179-182.
- (15) Hankinson S, Husebye E. Breast Cancer. In: Adami HO, Hunter D, Trichopoulos D, editors. *Textbook of Cancer Epidemiology*. Oxford University Press, 2002: 3001-339.
- (16) Tavani A, Braga C, La VC, Negri E, Russo A, Franceschi S. Attributable risks for breast cancer in Italy: education, family history and reproductive and hormonal factors. *Int J Cancer* 1997; 70(2):159-163.
- (17) Bakken K, Alsaker E, Eggen AE, Lund E. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. *Int J Cancer* 2004; 112(1):130-134.
- (18) Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. *Am J Epidemiol* 1998; 147(9):826-833.
- (19) Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med* 2002; 347(16):1227-1232.

- (20) Purushotham AD, Upponi S, Klevesath MB, Bobrow L, Millar K, Myles JP et al. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. *J Clin Oncol* 2005; 23(19):4312-4321.
- (21) Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer. Issue 1. Art. No.: CD000486. DOI: 10.1002/14651858.CD000486. 2001. Cochrane Database of Systematic Reviews. Ref Type: Report
- (22) Ecsedy J, Hunter D. The Origin of Cancer. In: Adami HO, Hunter D, Trichopoulos D, editors. *A Textbook of Cancer Epidemiology*. Oxford University Press, 2002: 29-53.
- (23) Perou CM, Sorlie T, Eisen MB, van de RM, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. *Nature* 2000; 406(6797):747-752.
- (24) Spratt JS, Spratt JA. Growth rates. In: Donegan WL, Spratt JS, editors. *Cancer of the Breast*. Saunders, 2002: 443-476.
- (25) Meyer JS, Prey MU, Babcock DS, McDivitt RW. Breast carcinoma cell kinetics, morphology, stage, and host characteristics. A thymidine labeling study. *Lab Invest* 1986; 54(1):41-51.
- (26) Spang-Thomsen M, Nielsen A, Visfeldt J. Growth curves of three human malignant tumors transplanted to nude mice. *Exp Cell Biol* 1980; 48(2):138-154.
- (27) Spratt JS, Meyer JS, Spratt JA. Rates of growth of human solid neoplasms: Part I. *J Surg Oncol* 1995; 60(2):137-146.
- (28) Spratt JS, Meyer JS, Spratt JA. Rates of growth of human neoplasms: Part II. *J Surg Oncol* 1996; 61(1):68-83.
- (29) Peer PG, van Dijck JA, Hendriks JH, Holland R, Verbeek AL. Age-dependent growth rate of primary breast cancer. *Cancer* 1993; 71(11):3547-3551.
- (30) von Fournier D, Weber E, Hoeffken W, Bauer M, Kubli F, Barth V. Growth rate of 147 mammary carcinomas. *Cancer* 1980; 45(8):2198-2207.
- (31) Friberg S, Mattson S. On the growth rates of human malignant tumors: implications for medical decision making. *J Surg Oncol* 1997; 65(4):284-297.
- (32) Adami HO, Signorello LB, Trichopoulos D. Towards an understanding of breast cancer etiology. *Semin Cancer Biol* 1998; 8(4):255-262.
- (33) Tabar L, Fagerberg G, Day NE, Duffy SW, Kitchin RM. Breast cancer treatment and natural history: new insights from results of screening. *Lancet* 1992; 339(8790):412-414.
- (34) Duffy SW, Tabar L, Fagerberg G, Gad A, Grontoft O, South MC et al. Breast screening, prognostic factors and survival--results from the Swedish two county study. *Br J Cancer* 1991; 64(6):1133-1138.
- (35) Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL et al. Randomized trial of breast self-examination in Shanghai: final results. *J Natl Cancer Inst* 2002; 94(19):1445-1457.
- (36) Morrison AS. Is self-examination effective in screening for breast cancer? *J Natl Cancer Inst* 1991; 83(4):226-227.
- (37) Andersson I. [Some glimpses from the history of mammography: from radiography of preparations to general screening]. *Lakartidningen* 1996; 93(14):1329-1333.

- (38) Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 1982; 69(2):349-355.
- (39) Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985; 1(8433):829-832.
- (40) Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ* 1988; 297(6654):943-948.
- (41) Breast cancer screening - National Cancer Institute. <http://www.cancer.gov> . 2007.
Ref Type: Electronic Citation
- (42) Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* 1995; 75(10):2507-2517.
- (43) Harris JR, Lippman ME, Veronesi U, Willett W. Breast cancer (1). *N Engl J Med* 1992; 327(5):319-328.
- (44) World Health Organization. IARC Handbooks of Cancer Prevention: Handbook 7: Breast Cancer Screening. IARC Press, 2001.
- (45) Olsen O, Gotzsche PC. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2001;(4):CD001877.
- (46) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002; 359(9310):909-919.
- (47) Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2006;No.: CD001877. DOI: 10.1002/14651858.CD001877.pub2.
- (48) Press release; Mammography screening can reduce deaths from breast cancer. www.iarc.fr . 2007. 2007.
Ref Type: Electronic Citation
- (49) NCI Statement on Mammography Screening. National Cancer Institute webpage . 23-3-2007.
Ref Type: Electronic Citation
- (50) Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB et al. Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ* 2005; 330(7485):220.
- (51) Schmidt JG. Natural history of breast cancer. *Lancet* 1992; 339(8796):810.
- (52) Moller B, Fekjaer H, Hakulinen T, Sigvaldason H, Storm HH, Talback M et al. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Stat Med* 2003; 22(17):2751-2766.
- (53) 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(12):39-45.
- (54) Cancer Registry of Norway. Cancer in Norway 2005. <http://www.kreftregisteret.no> . 2007. 1-6-2007.
Ref Type: Electronic Citation
- (55) Hofvind S. The Norwegian Breast Cancer Screening program: Selected Process Indicators and their Utilization in Epidemiological Research. University of Oslo, Norway, 2005.

- (56) Insightful. S-PLUS. <http://www.insightful.com> . 2002.
Ref Type: Electronic Citation
- (57) R core group. R. <http://www.r-project.org> . 2003. 11-9-2006.
Ref Type: Electronic Citation
- (58) Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 1987; 6(4):469-481.
- (59) Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med* 1987; 6(4):449-467.
- (60) Karlin S, Taylor H. *An Introduction to Stochastic Modeling*. Rev. ed. ed. Academic Press, 1994.
- (61) Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984; 40(1):1-14.
- (62) Chen HH, Duffy SW, Tabár L. A mover-stayer mixture of Markov chain models for the assessment of dedifferentiation and tumour progression in breast cancer. *Journal of Applied Statistics* 1997; 24(3):265-278.
- (63) van Oortmarssen GJ, Habbema JD, van der Maas PJ, de Koning HJ, Collette HJ, Verbeek AL et al. A model for breast cancer screening. *Cancer* 1990; 66(7):1601-1612.
- (64) Chen HH, Duffy SW, Tabár L, Day NE. Markov Chain models for progression of breast cancer Part I: tumour attributes and the preclinical screen-detectable phase. *Journal of Epidemiology and Biostatistics* 1997; 2(1):9-23.
- (65) Chen HH, Prey MU, Babcock DS, Day NE. Breast carcinoma cell kinetics, morphology, stage, and the preclinical screen-detectable phase. *Lab Invest* 1997; 2(1):9-23.
- (66) Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Stat Med* 2002; 21(22):3369-3382.
- (67) 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. *Stat Med* 1995; 14(14):1531-1543.
- (68) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed* 1985; 20(1):79-93.
- (69) CISNET Home Page. CISNET webpage . 23-3-2007.
Ref Type: Electronic Citation
- (70) Croyle R. Guest Update by Dr. Robert Croyle; CISNET Offers Powerful New Tools for Cancer Control. National Cancer Institute webpage . 2007.
Ref Type: Electronic Citation
- (71) CISNET Home Page; Model Profiles; University of Wisconsin. CISNET webpage . 23-3-2007.
Ref Type: Electronic Citation
- (72) Spratt JA, von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. *Cancer* 1993; 71(6):2013-2019.
- (73) 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(8):1506-1510.

- (74) 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). *Int J Epidemiol* 1991; 20(4):852-858.
- (75) Adami HO. Long-term consequences of estrogen and estrogen-progestin replacement. *Cancer Causes Control* 1992; 3(1):83-90.
- (76) Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002; 288(3):321-333.
- (77) 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 Obstet Gynecol Scand* 2004; 83(9):850-856.
- (78) Bakken K, Alsaker E, Eggen AE, Lund E. [Estrogen replacement therapy and breast cancer]. *Tidsskr Nor Laegeforen* 2005; 125(3):282-285.
- (79) Legemiddel forbruk i Norge [Medical drugs used in Norway - website in Norwegian]. <http://www.legemiddelforbruk.no/> . 2007.
Ref Type: Electronic Citation
- (80) Jonsson H, Johansson R, Lenner P. Increased incidence of invasive breast cancer after the introduction of service screening with mammography in Sweden. *Int J Cancer* 2005; 117(5):842-847.
- (81) Lund E, Bakken K, Dumeaux V, Andersen V, Kumle M. Hormone replacement therapy and breast cancer in former users of oral contraceptives-The Norwegian Women and Cancer study. *Int J Cancer* 2007.
- (82) Hofvind S, Moller B, Thoresen S, Ursin G. Use of hormone therapy and risk of breast cancer detected at screening and between mammographic screens. *Int J Cancer* 2006; 118(12):3112-3117.
- (83) Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* 2007; 356(16):1670-1674.
- (84) van Oortmarssen GJ, Habbema JD, Lubbe JT, van der Maas PJ. A model-based analysis of the HIP project for breast cancer screening. *Int J Cancer* 1990; 46(2):207-213.
- (85) Perry N, Broeders M, de Wolf C, Törnberg S. European Guidelines for Quality Assurance in Mammography Screening. 3 ed. ed. European Commission, 2001.
- (86) Haus AG. Historical technical developments in mammography. *Technol Cancer Res Treat* 2002; 1(2):119-126.
- (87) 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 Oncol* 2004; 43(5):467-473.
- (88) Lynge E, Olsen AH, Fracheboud J, Patnick J. Reporting of performance indicators of mammography screening in Europe. *Eur J Cancer Prev* 2003; 12(3):213-222.
- (89) Porter PL, El Bastawissi AY, Mandelson MT, Lin MG, Khalid N, Watney EA et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 1999; 91(23):2020-2028.
- (90) Kavanagh AM, Mitchell H, Giles GG. Hormone replacement therapy and accuracy of mammographic screening. *Lancet* 2000; 355(9200):270-274.
- (91) Litherland JC, Stallard S, Hole D, Cordiner C. The effect of hormone replacement therapy on the sensitivity of screening mammograms. *Clin Radiol* 1999; 54(5):285-288.

8 Appendix: Quantifying the level of unregistered mammography screening

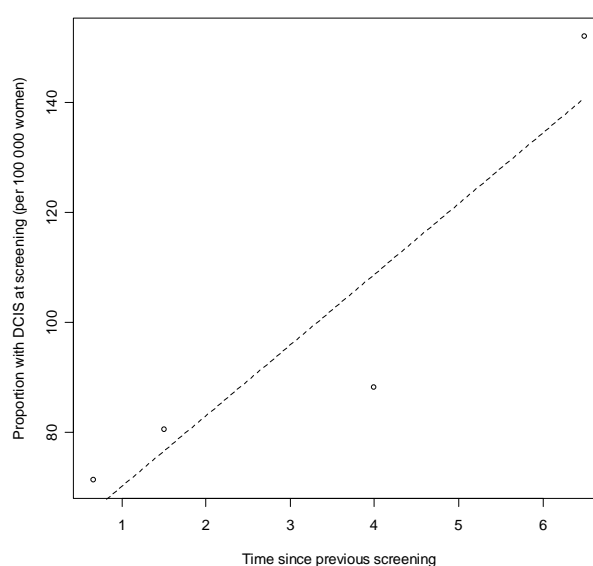
As we have no possibility to separate “true interval cancers” from cases found on opportunistic screening, our estimates in papers II-IV may be biased by opportunistic screening. Inquiries by the Cancer Registry of Norway have indicated that approximately 10 % of women in the NBCSP target group are screened each year, but these estimates are uncertain and a more direct way of estimating the level of opportunistic screening would give additional information.

Usually, non-invasive DCIS cancers are mainly found on mammography screening as most DCIS cancers give few clinical symptoms. Still, 8.6 % of interval cancers in the NBCSP are DCIS cancers, indicating that a substantial proportion of the registered interval cancers arise from opportunistic screening. Assuming a given proportion of DCIS cancers among clinically detected cancers, the volume of DCIS cases found on opportunistic screening can be estimated. Combining the volume of DCIS cases found at opportunistic screening with the observed proportions of DCIS cancers found in women with different time since previous screening, it is possible to estimate the expect number of opportunistic screenings needed to find the given DCIS level.

Applying the questionnaire data on time since previous screening used in paper III, we estimated the level of opportunistic screening by:

- 1) Estimating $P(\text{DCIS found} \mid X \text{ months since last screening})$, using linear regression on the proportion of DCIS cases as a function of time since last screening (Figure 17)

Figure 17: Linear regression of the proportion of cases with DCIS as function of time since previous screening



- 2) Estimating the probability that a woman have one opportunistic screening T months after screening by:

$$P(\text{opportunistic screening } T \text{ months after last NBCSP visit}) = \frac{[\text{DCIS}_O(T) - \text{DCIS}_E] \cdot N(T)}{P(\text{DCIS found} \mid X \text{ months since last screening})}$$

where:

$\text{DCIS}_O(T)$ = The observed proportion of DCIS cases at time T

DCIS_E = The expected proportion of DCIS cases without screening

$N(T)$ = The number of interval tumours (DCIS + invasive cancers)

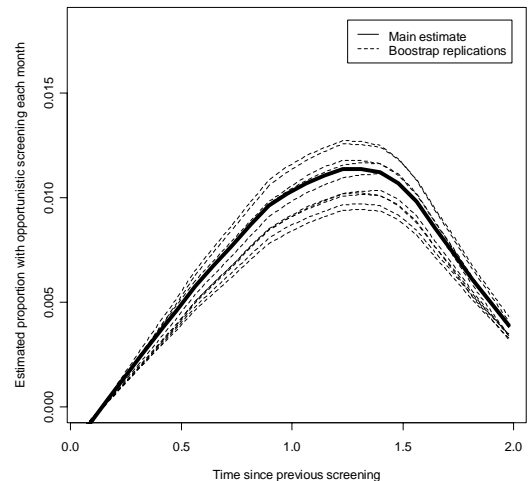
Using this new approach on the NBCSP dataset, we find relatively stable estimates given different assumed DCIS proportions without screening (Table 5). Assuming 1 % DCIS in clinical dataset without opportunistic screening, the proportion of women having one opportunistic screening in the interval was estimated to be 17.9 %, with a 95% bootstrap confidence interval of { 14.4 , 23.7 }.

These estimates are, of course, uncertain. A key assumption is the proportion of DCIS without screening, but there are also other assumptions of importance. As we combine data concerning private and public screening examinations, we assume that there is no fundamental difference in the detection of DCIS cases between private and public screening institutes. In addition, we assume that each woman has maximum one opportunistic screening in the interval between the two subsequent NBCSP screening examinations. All these

Table 5: Estimated proportion of women having one opportunistic screening in the two year interval following screening, assuming different DCIS proportions among cases found clinically

Assumed DCIS proportion without screening (%)	Estimated proportion of women having one opportunistic screening in the two year interval (%)
0.0	20.1
0.5	19.0
1.0	17.9
2.0	15.6

Figure 18: Smoothed estimated frequencies of opportunistic screening with time since previous screening



assumptions can be discussed, but there are also several indications for assuming that the DCIS proportion is a good proxy for opportunistic screening:

- 1) The estimated overall level of opportunistic screening fits well with information from earlier inquiries done by the Cancer Registry of Norway.
- 2) There is a distinct peak in the estimated proportion of women going to opportunistic screening between 1 and 1.5 years after the first NBCSP attendance (Figure 18). This corresponds well with many women going to opportunistic screening in the mid period between two public screening examinations, with some additional delay from the examination to final diagnosis.
- 3) Including data on women not attending the next public screening examination, we find a sharp increase in DCIS rates shortly after the invitation to the next screening attendance (data not shown), fitting well with the assumption that some women chose to attend private screening instead of the repeated public screening.

Overall, there are substantial uncertainties regarding the level of opportunistic screening, but this new approach gives, at least, an indication of the true level of opportunistic screening and bias of the estimated values in papers II, III and IV. Without any better method available, we have chosen to include these estimates in the discussion to illustrate the problem of biased estimates due to opportunistic screening.

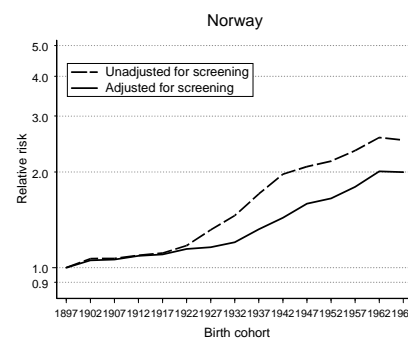
Summary for the general public (in Norwegian)

Brystkreft er den vanligste årsaken til tapte leveår hos Norske kvinner under 65 år, og har dessverre vist seg å være relativt vanskelig å forebygge. For å bedre overlevelsen, innførte Norge gradvis et offentlig mammografiscreening program i 1996, som ble landsdekkende i 2004, med mål om å bedre overlevelsen via tidligere diagnose/behandling.

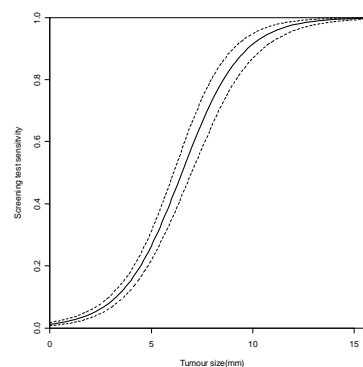
Selv om studier tyder på en gevinst ved mammografiscreening, er det fortsatt mye uklart når det gjelder hvor ofte og til hvilke aldersgrupper tilbudet bør gies. Sentrale spørsmål i denne sammenheng er hvor fort brystkreftsvulster vokser, og hvor tidlig i utviklingen mammografiscreeningen klarer å avdekke brystkreft. Dessverre har informasjon om veksthastigheten til brystkreftsvulster stort sett vært basert på små og selekterte kliniske studier, da så godt som all diagnostisert brystkreft i land med god registrering behandles.

Når mammografiscreeningen starter avdekkes nye brystkreft tilfeller, hvorpå man de påfølgende årene ser færre tilfeller siden, ettersom tilfeller som normalt ville oppstått på ett senere tidspunkt allerede er avdekket på screeningen. Som et alternativ til kliniske studier kan vekstraten estimeres ved å utnytte disse variasjonene i brystkreft hyppighet. I dette arbeidet har vi studert disse variasjonene, og bygd matematiske modeller for å kunne anslå både den underliggende veksthastigheten til brystkreftsvulstene og sensitiviteten til screening undersøkelsene.

I det første arbeidet ser vi at risikoen for brystkreft har økt betydelig i Norge, selv korrigert for økt screening. I arbeide nummer to og tre ser vi at det norske screeningprogrammet trolig flytter diagnosen lengre enn hva som er anslått for mange tidligere programmer, men også med betydelig høyere usikkerhet. I det siste arbeidet går vi videre og bygger en helt ny modell for å estimere veksthastigheten til brystkreftsvulster. Resultatene tyder på svært store individuelle variasjoner i veksthastighet, hvor noen svulster bruker under en måned på å vokse fra 10 til 20 mm i diameter, mens andre bruker over 6 år. Disse resultatene utnyttet så til å vise hvorfor brystkreft screening i høy alder (70+) trolig kan gi ett betydelig antall unødige diagnoser, ettersom screeningen kan avdekke tilfeller som normalt ikke ville oppstått innen kvinnens gjenværende levetid.



Relative risk of breast cancer for Norwegian women born between 1897 and 1967, as estimated in paper I, with and without correction for screening



Screening test sensitivity in the Norwegian breast cancer screening program, as estimated in paper IV

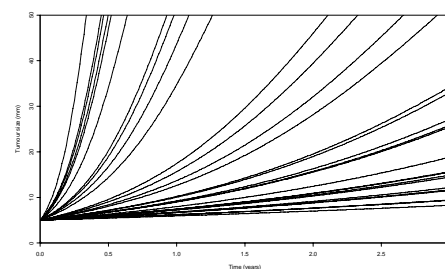


Illustration of the estimated large variation in breast cancer tumour growth rates; Thirty random growth rates based on the values estimated in paper IV