NTNU

Norwegian University of
Science and Technology

A simulation of the development and screening of Cancer Mammae

Marianne Riksheim

Master of Science in Statistics

Submission date: May 2009

Supervisor: Bo Henry Lindqvist, MATH

Co-supervisor: Harald Weedon-Fekjær, Kreftregisteret

Lars Vatten, Institutt for samfunnsmedisin

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Based on the tumor growth model of Harald Weedon-Fekjær et al.'s paper "Breast cancer tumor growth estimated through mammography screening data", a simulation of breast cancer occurrence and tumor growth in a large population of women was made. The simulation was made realistic by starting tumor growth according to a Poisson process, including a distribution for clinical detection size and a screening test sensitivity function, and using an individual growth rate, κ, based on estimates of Weedon-Fekjær et al. [1]

After running the full simulation, parts of the simulation outcomes were compared to known data of breast cancer and the model was found to give realistic and expected results. The simulation was then used to look for other interesting results such as expected reduction in time to tumor detection due to screening and finding the size distribution of tumors before and after screening.

For the age group 50 - 69 years, it was found that screening every year allows a reduction of 19.1 months, while screening every two, three, five and ten years allows for reductions of, respectively, 9.5, 8.3, 6.1 and 3.4 months. These results are based on the assumption that tumors are actually found at screening, i.e. the tumors are not found clinically before they are found on screening.

When clinical findings are included, different results are obtained. For the age group 50 - 69 years, it was found that screening every year allows a reduction of 15 months, while screening every two, three, five and ten years allows for reductions of, respectively, 4.8, 1.4, -5.6 and -21.5 months. Negative numbers indicate that a tumor is found earlier clinically than at screening.

To my grandmother Wilhelmina, looking down at me so proud.

Contents

Main introduction

Based on data from 2002, 10.9 million people worldwide are diagnosed with cancer every year and there are 6.7 million deaths from cancer [2]. One in ten of all new cancers diagnosed, and close to one in four cancers diagnosed in women worldwide, is a cancer of the breast. More than 1.1 million women are diagnosed each year and breast cancer is the main cause of death from cancer in women globally [3].

The Cancer Registry of Norway (CRN) is one of the oldest national cancer registries in the world. By law, all new cancer cases must be notified to the CRN, which makes the registry unique. The Norwegian Breast Cancer Screening Program (NBCSP) is one of the two screening programs organized by the CRN. The NBCSP obtained national coverage in 2004. The program invites all women between 50 and 69 years of age to mammography screening every two years. The idea behind breast cancer screening, is to detect a tumor earlier than it would have been without screening and therefore be able to diagnose the patient at an earlier stage. The effects of screening is a topic still undergoing discussion.

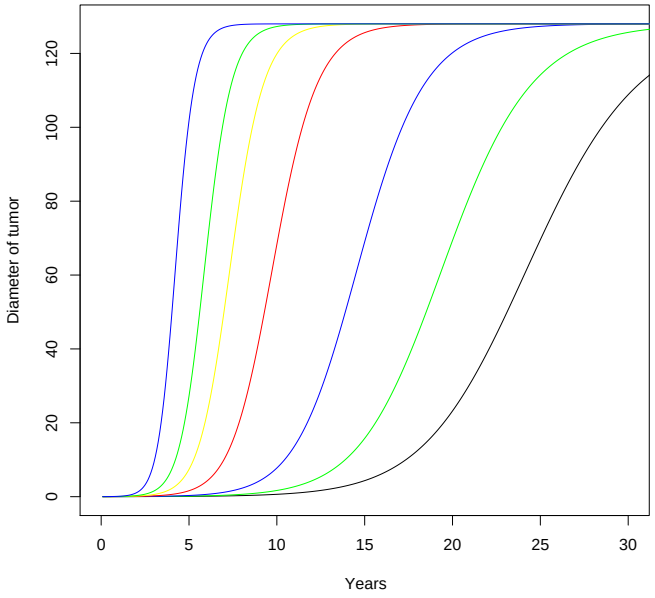


Figure 1: The full path of seven breast cancer tumor growth curves, with different growth rates

In Harald Weedon-Fekjær et al.'s [1] article "Estimating tumour growth and screening test sensitivity", a model for breast cancer tumor growth was presented. The model is based on previous work by Spratt et al. [4], among others. Weedon-Fekjær presents a tumor growth function including an individual growth rate, which makes it possible to simulate the full path of a breast cancer tumor's development. An example of seven growth curves, each with a different growth rate, is given in figure 1. This is the point departure for this thesis.

In this thesis, a synthetic population based on the estimated model will be studied. This has many advantages, as things like screening frequencies and population characteristics easily can be varied. Also, when or how a tumor is detected can be changed. These modifications give a unique possibility to investigate matters that are usually not available for research from real data, for example how much earlier a tumor could be detected with screening than without.

The first part of this thesis consists of a theoretical background. Different aspects of cancer mammae (breast cancer) are provided, previous work on modeling of breast cancer and different parameters needed in the models are presented and a few statistical methods are described.

The second part gives a short presentation of the statistical programming language R and an overview of how the simulation works.

In part three, the results of the simulation are presented, the results are discussed and the conclusion is given.

The last part is an appendix containing some of the R-code used in the simulation.

Part I
Theory

Introduction

The goal of this part, is to describe the theory behind the steps needed to make a realistic simulation of cancer mammae in a Norwegian female population. A short introduction of cancer mammae will be given, models explained and statistical methods presented.

1 Cancer mammae

1.1 What is cancer?

Cancer occurs in the body when cells start growing without regulation and the tumor created invades surrounding tissue. When cells grow unrestrained, they are often called cancer cells and a tumor consisting of cancer cells, is called a malignant tumor.

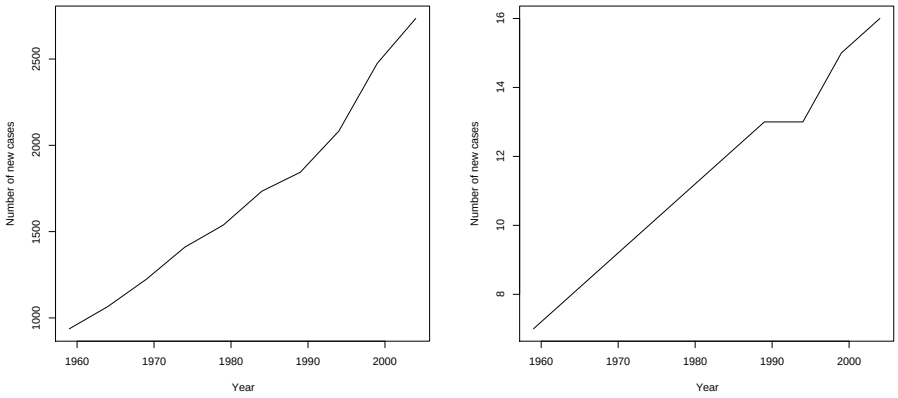
1.2 Cancer mammae

Cancer mammae is cancer of the breast, and in daily language mostly referred to as breast cancer. The breast consists of fatty-, glandular- and connective tissue. In women, the breast also contains mammary glands (milk glands) and lactiferous ducts (milk ducts). About 80 % of all cancer mammae cases, originate in the lactiferous ducts, while 10-15 % of the cases occur in the mammary glands [5]. This difference in anatomy is one of the reasons cancer mammae is by far more common in women than in men. In women, cancer mammae is the most common type of cancer.

Cancer cells that emerge in the breast tissue, mostly originate in the breast it self; spreading of cancer cells from other organs to the breast is very rare. On the other hand, if a malignant tumor in the breast grows into a blood- or lymphatic vessel, cancer cells may unfasten and be carried with the blood- or lymph stream to other locations in the body, called metastasis.

Cancer mammae is characterized by a great variation in progression; from fast growing tumors that spread to other organs (metastasis) to slow growing tumors without metastasis. In 25-30% of the cases, the tumor is aggressive [6].

Experience shows that younger women, on average, have more aggressive forms of cancer mammae then older women.



(a) Women (b) Men

Figure 2: Average number of annual new cases of cancer mammae in a) women and b) men from 1957-2006

Table 1: Average number of annual new cancer mammae cases in Norwegian men and women from 1957 - 2006, in five years intervals

	1957-61	1962-66	1967-71	1972-76	1977-81
Men	7	8	9	10	11
Women	937	1066	1224	1410	1538
	1982-86	1987-91	1992-96	1997-01	2002-06
Men	12	13	13	15	16
Women	1734	1844	2082	2475	2735

1.3 History of cancer mammae

According to the report "Cancer in Norway 2006" [7] there were 2687 new cases of cancer mammae in Norway in 2006, where 2673 cases were in women and 14 in men.

Table 1 and Figure 2 show the number of new cases of cancer mammae in Norway the last 50 years. As seen, there has been a strong increase of new cases.

Table 2 shows the number of annual new cases the last ten years. Comparing table 1 and Table 2, it can be seen that although there has still been an increasing trend the last ten years, it has not been as strong as earlier. In 2006 there was even a decrease in number of annual new cases for the first time in many years.

Table 2: Number of annual new cancer mammae cases in Norwegian men and women from 1997 - 2006

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Men	11	21	14	17	13	14	20	14	18	14
Women	2401	2416	2408	2528	2621	2694	2723	2787	2799	2673

1.4 Risk factors

Some known factors that increase the risk of cancer mammae are:

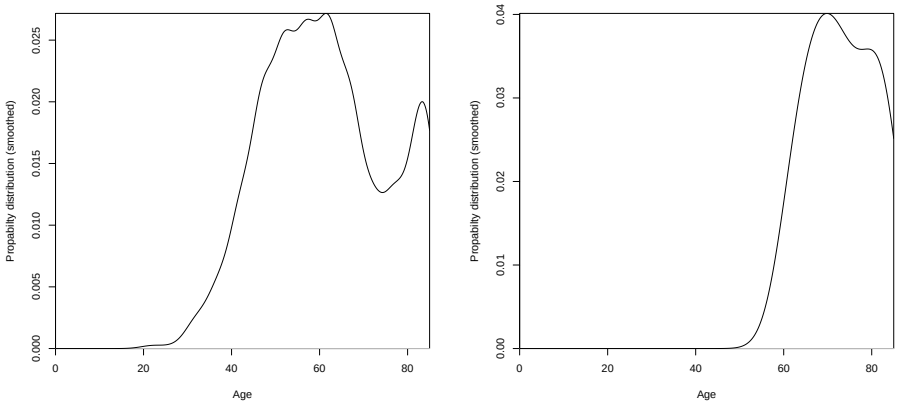
- Sex (being female)
- Increasing Age
- Inheritance (Close relatives; mother, daughter, sister and grandmother)
- Hormonal conditions (Early menstruation, late first birth, childlessness and late menopause)
- Case history (Cancer in one breast, increases chance of cancer in the other breast)
- Long time use of estrogen therapy before menopause
- Excessive weight after menopause, especially for tall women
- Alcohol.

Behind sex, the most prominent risk factor for developing cancer mammae is increasing age. Table 3 and Figure 3 show the age distribution of annual new cases in 2006. At a young age, the probability of developing cancer mammae is fairly low, but risk increases rapidly from age 40-44 in women, and somewhat later in men, to the highest risk at age 60-64 in women.

An interesting finding is the drop in number of annual new cases around age 70. The drop is a result of women leaving the official screening program which runs from age 50 to 69. A new drop can be seen at age 80+. This drop can be explained by the lack of detection of disease in elderly, as they often decease from other causes.

Table 3: The age distribution of new cases of cancer mammae in 2006

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
Men	0	0	0	0	0	0	0	0	0
Women	0	0	0	0	4	3	36	75	167
	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
Men	0	0	0	2	3	3	2	3	1
Women	295	345	351	369	290	163	175	195	205



(a) Women (b) Men

Figure 3: The distribution of new cancer mammae cases in different age groups

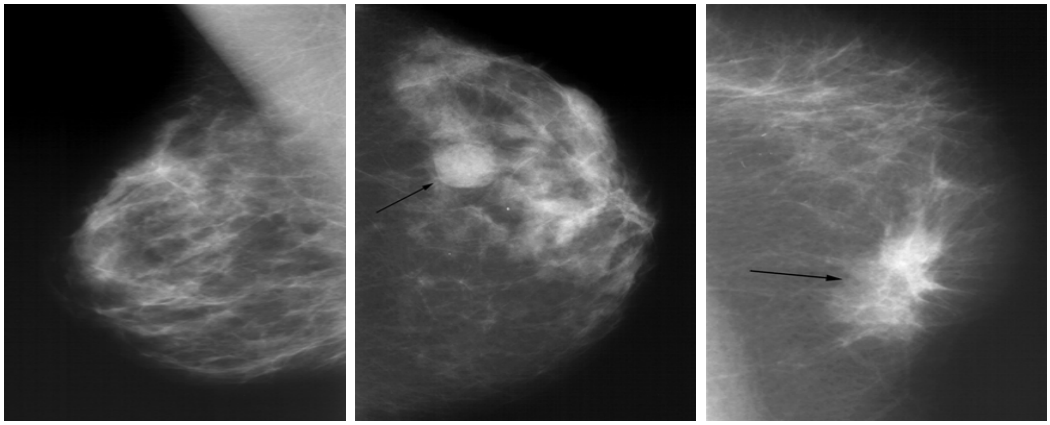
1.5 Risk reduction

Similar to factors that increase the risk of cancer mammae, there are also factors that may reduce the risk:

- First pregnancy at young age ($<$ 20-25 years)
- Several births at young age ($<$ 25 years)
- Asian descent
- Regular exercise.

1.6 Signs and symptoms

Usually, the first and most prominent symptom of cancer mammae is a tumor, which can be felt as a lump in the breast or armpit. Other symptoms may include chest pain, fluid from the nipple, changes in the nipple, skin change and open wounds, but these are less common. In cases that have existed for a long time before diagnosis, symptoms of back- and chest pain are sometimes signs of metastasis. Shortness of breath could indicate metastasis to the lungs and nausea and stomach pain may be a sign of metastasis to the liver.



(a) Normal (b) Benign (c) Malignant

Figure 4: Mammography pictures of breasts without a tumor, with a benign tumor and a malignant tumor. Source: http://www.pasienthandboka.no/

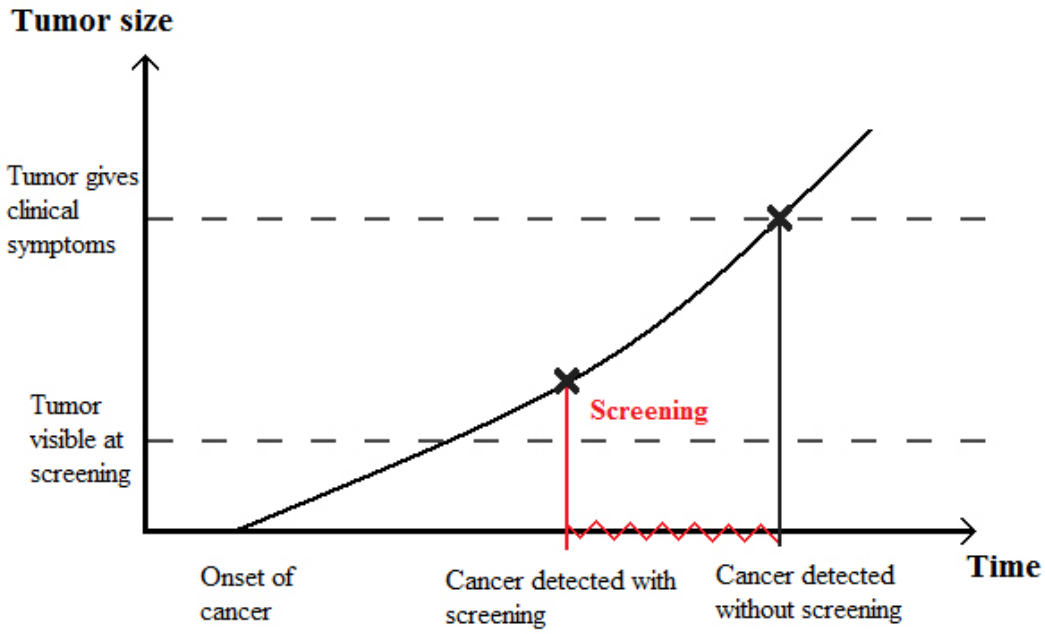


Figure 5: The idea behind screening

1.7 Screening by mammography

Mammography is a x-ray photo of the breast. Figure 4 shows mammography images of a normal breast 4(a), a breast with a benign tumor 4(b) and a breast with a malignant tumor 4(c).

Screening is a term used to describe different types of examinations that are performed on a large number of individuals. Diseases that are screened for, are often rare, but if found at an early stage, treatment is thought to be more efficient.

Figure 5 shows the idea of mammography screening for cancer mammae. At any given time, a tumor might start growing, marked 'onset of cancer'. The tumor will start growing, slow or fast, depending on factors such as age. Reaching a certain size, the tumor is large enough to be visible on mammography. In most cases, this occurs before clinical signs can be detected, and before any symptoms are present.

So, the idea behind screening by mammography, is to detect the tumor when it is large enough to be found at mammography, but before the tumor gives clinically detectable symptoms. The time spent in this phase is called sojourn time.

The overall mean sojourn time (MST) has been estimated by Weedon-Fekjær et al. [1] to be 2,9 years among all cases and significantly longer in older women. Earlier diagnosis opens for earlier treatment which could give, a better prognosis.

Figure 6 shows the size distribution of tumors found at first screening for women reporting no earlier screening history.

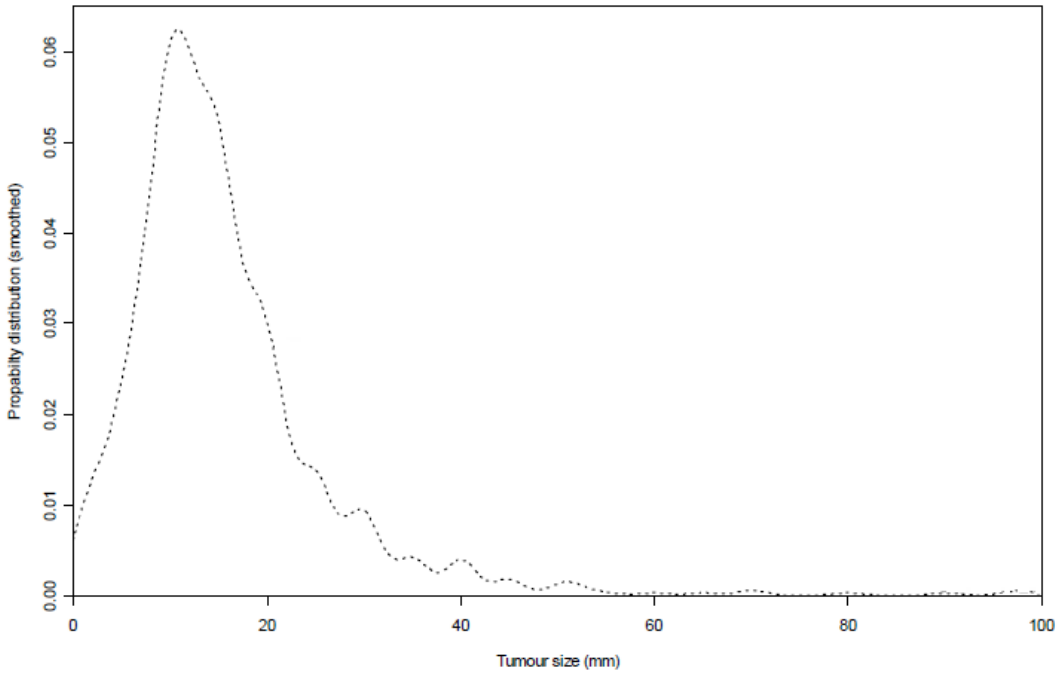


Figure 6: The size distribution of tumor diameters found at screening. Data from the Norwegian Breast Cancer Screening Program (1995-2002). Figure from Weedon-Fekjær et al. [1]

1.8 Other diagnostics and treatment

Diagnostics of cancer mammae is not done by mammography screening alone. Clinical- and self-examinations are also important diagnostic tools. When a lump or irregularity is found, either by mammography or examination, additional inspection is needed. Ultrasound is an important supplement to mammography and in special cases magnetic tomography (MR) is used. If the preliminary tests indicate a malignant tumor, a needle biopsy is performed. A final cancer diagnosis can only be given by examination of a tissue sample under a microscope.

The treatment of cancer mammae depends on signs of metastases. Local tumors without signs of metastases can be removed by surgery alone, with good effect. In other cases treatment, alone or in combination, consist of radiation therapy, hormone therapy and chemotherapy.

1.9 The Cancer Registry of Norway and the Norwegian Breast Cancer Screening Program

The Cancer Registry of Norway (CRN) was established in 1951. It is one of the oldest national cancer registries in the world. By law, all new cancer cases must be notified to the CRN. This makes the registry unique.

The Norwegian Breast Cancer Screening Program (NBCSP) is one of the two screening programs organized by the CRN. The NBCSP started in four Norwegian counties in 1995 and was gradually introduced to the rest of the counties until full coverage was obtained in 2004. The program invites all women between 50 and 69 years of age to mammography (see section 1.7) every two years.

1.10 Hormone replacement therapy and NBCSP data collection

Hormone replacement therapy (HRT) is a drug used to reduce physical disturbances associated with menopause in women. Early research showed a protective effect of HRT use, especially against cardiovascular disease, but also against other severe health problems. Because of the increased belief in a protective effect, HRT use increased rapidly until the late 1990s. In 2002 the Women's Health Initiative study reported an increased risk of both breast cancer and cardiovascular disease in a large randomized trial [8] and HRT use dropped.

Figure 7 shows annual Norwegian sales of HRT drugs. As indicated, the NBCSP data used to estimate the parameters in Weedon-Fekjær et al.'s [1] model, were collected at the same time HRT use was at its peak.

Weedon-Fekjær et al. corrected their data for the extended use of HRT-drugs at the time of data collection.

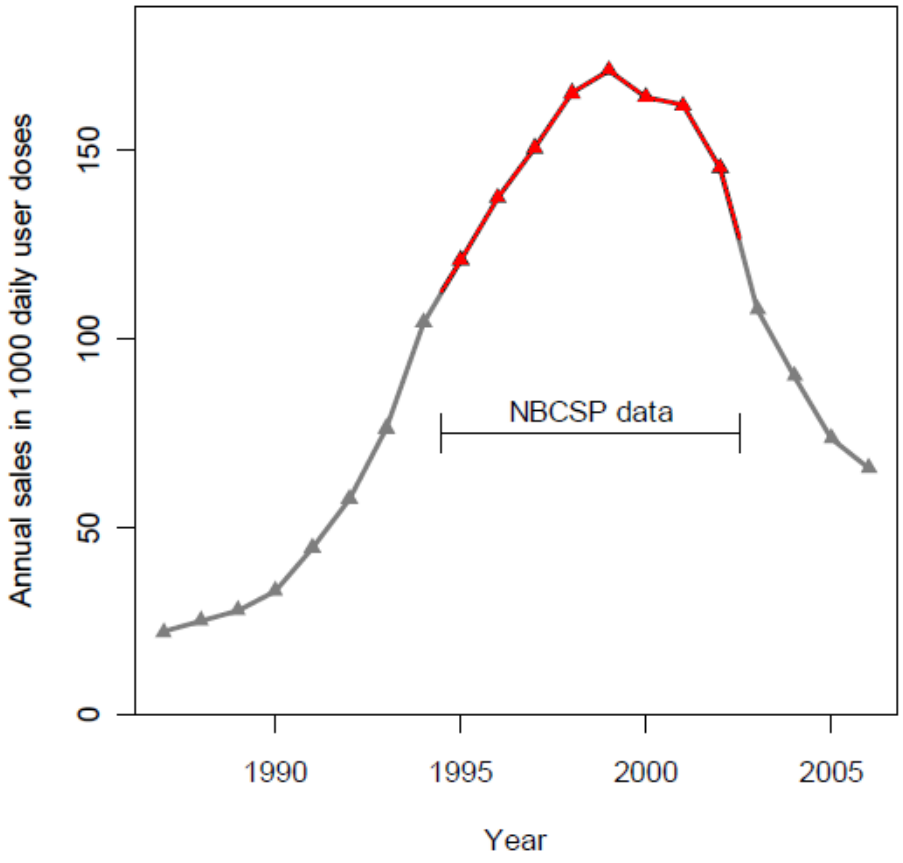


Figure 7: Use of hormone replacement therapy (HRT) in Norway from 1987 to 2006, in daily doses. The collection of data used to estimate the Weedon-Fekjær et al.'s [1] model parameters, were collected from 1995 to 2002. Figure from Weedon-Fekjær et al. [1]

2 Previous work

2.1 Models

Since cancer mammae is one of the most common types of cancer in women, especially in the western world, several attempts have been made to simulate the growth of breast cancer tumors. Looking back to Figure 5, knowledge of the growth of these tumors is interesting because it can help us determine when a tumor is in the scrambled area; large enough to be detected at screening by mammography, but before it gives clinical symptoms.

Modeling of cancer mammae tumor growth has been done in different ways. Two main methods have been used in the past; a Markov chain model with appurtenant transition probabilities, and assigning a distribution to the size of the tumor as a function of time. Also, a computer simulation program, MISCAN, has been used to model cancer mammae.

The Markov model One way to define a basic Markov model of screening related cancer progression, is a three state model thoroughly studied by S. W. Duffy et al. [9] - [11]. The different states of a basic Markov model are shown in Figure 8.

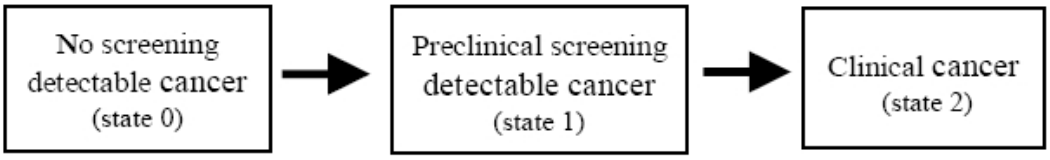


Figure 8: The three states of a basic Markov model. Figure from Weedon-Fekjær et al. [1]

MISCAN-Fadia MISCAN [12] is a computer simulation program developed for building models for cancer screening. The models are used to analyze and explain results of cancer screening trials, to predict and compare the (cost-) effectiveness of different screening policies, and to monitor the results of population screening programs.

The MISCAN-Fadia model includes the concept of a "fatal diameter" of cancer mammae tumors. The model has been used to model survival and screening benefits and is based on continuous tumor growth. It consists of four major components: population, natural history, screening, and treatment, where the natural history component is based on a cohort version of the MISCAN-Fadia model. This is illustrated in Figure 9.

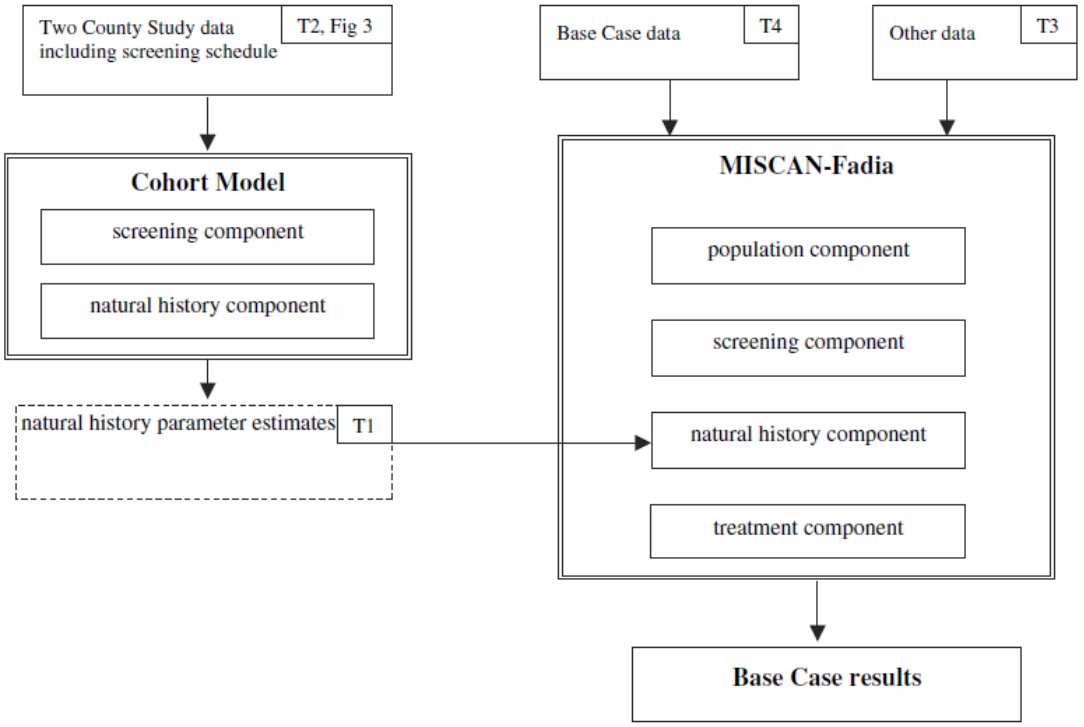


Figure 9: The two MISCAN-Fadia simulation models used for producing the base case results. Figure from Sita Y. G. L. Tan et al. [12]

2.2 Assigning a distribution

The growth rate in cancer mammae tumors is very individual. This makes assigning one specific model to the growth hard. A much used model is that of exponential growth. Since a growing tumor, slow or fast, is at some point restricted by conditions in the host (immunologic response, vascular supply etc.), it is natural to consider models that "decelerate". The Gompertz and logistic distribution satisfy this condition.

Spratt et. al. [4] investigate these three distributions, with different parameters, for their fit to cancer mammae tumor growth data.

Exponential The exponential model suggests that a tumor increases in size at a constant rate without any growth deceleration until death of the host. This model gives the worst fit to the data tested by Spratt et. al.

$$V(t) = V_{cell} \cdot e^{t\kappa} \ , \tag{1}$$

where $V(t)$ is the tumor volume at time t, κ is the relative growth rate and V_{cell} is the volume of one cell, i.e. $V_{cell} = V(0)$.

Gompertz The next model considered is the Gompertz model. The general form is

$$V(t) = V_{cell} \cdot e^{(1-e^{-\alpha t})\beta/\alpha} \ , \tag{2}$$

where α and β are parameters describing the rate of decay of the relative growth rate and the relative growth rate it self.

It can be shown that:

$$V_{max} = \lim_{t\to\infty} V(t) = V_{cell} \cdot e^{\left(\frac{\beta}{\alpha}\right)} \ .$$

Inserting this into equation 2:

$$V(t) = V_{max} \cdot e^{\left[e^{(-\alpha t)} \cdot \left(\frac{\beta}{\alpha}\right)\right]} \ .$$

V_{max} is estimated to be 2^{40} cells and the volume of one cell is estimated to be $1 \cdot 10^{-6}$ mm^3 by Spratt et al. [4].

This model demonstrates decelerating growth, and gives a decent fit to the data.

Generalized logistic In the logistic model, tumor growth for small tumors is nearly exponential, but "breaks" the growth as the tumors becomes larger.

The generalized model satisfies:

$$\frac{d(V(t))}{dt} = \kappa(t) \cdot V(t) \cdot g(t, V(t)) \ ,$$

where $\kappa(t)$ is the relative growth rate and g(t, V(t)) is a size dependent interaction term.

Several forms of this equation have been studied. Skekhan [13] used the following approach. Let

$$g(t, V(t)) = 1 - \left(\frac{V(t)}{V_{max}}\right)^N \ .$$

As the volume of the tumor $V(t)$ increases, $1 - \left(\frac{V(t)}{V_{max}}\right)^N$ decreases and the tumor growth is restricted. Assuming the relative growth rate, $\kappa(t)$, and the environmental carrying capacity (which is the maximum tumor size), V_{max}, to be constant, the following form of the logistic equation results:

$$\frac{dV(t)}{dt} = V(t) \cdot \kappa \cdot \left(1 - \left(\frac{V(t)}{V_{max}}\right)^N\right) \ .$$

The solution to this equation for $N > 0$ is

$$V(t) = V_{max} \cdot \left[1 + e^{(-N(\kappa t + c))}\right]^{\left(-\frac{1}{N}\right)} ,$$

where

$$c = \left(-\frac{1}{N}\right) \ln \left[\left(\frac{V_{max}}{V_{cell}}\right)^N - 1\right] .$$

Inserting c into V(t) we get:

$$V(t) = V_{max} \left(1 + e^{-N\left(\kappa \cdot t - \left(\frac{1}{N}\right) \ln\left[\left(\frac{V_{max}}{V_{cell}}\right)^N - 1\right]\right)}\right)^{\left(-\frac{1}{N}\right)}$$

$$= \frac{V_{max}}{(1 + [(\frac{V_{max}}{V_{cell}})^N - 1]\, e^{-N\kappa t})^{\frac{1}{N}}} . \tag{3}$$

Equation 3 with $N = 4$ is shown by Spratt et. al. [4] to give the best fit to cancer mammae tumor growth data.

This is the form used by Weedon-Fekjær et. al. [1] and which will be used further in this thesis.

2.3 Diameter of a tumor

When a tumor is detected and removed, either after clinical or mammography detection, its size is measured as a diameter, assuming it is ball shaped. From equation 3 we know how the volume of a cancer mammae tumor develops and we know the formula for the volume of a ball

$$V(t) = \frac{4}{3}\pi \left[\frac{X(t)}{2}\right]^3$$

where $X(t)$ is the diameter of a tumor at time t.

Because volume is a measurement that is hard to visualize, it may be easier to conceive the size of a tumor if given the diameter.

Solving for $X(t)$:

$$\frac{4}{3}\pi\left[\frac{X(t)}{2}\right]^3 = \frac{V_{max}}{(1+[(\frac{V_{max}}{V_{cell}})^{\frac{1}{4}}-1]e^{-\frac{1}{4}\kappa t})^4}$$

$$\left[\frac{X(t)}{2}\right]^3 = \frac{3\cdot V_{max}}{4\pi\cdot(1+[(\frac{V_{max}}{V_{cell}})^{\frac{1}{4}}-1]e^{-\frac{1}{4}\kappa t})^4}$$

$$X(t) = \sqrt[3]{\frac{6\cdot V_{max}}{\pi\cdot(1+[(\frac{V_{max}}{V_{cell}})^{\frac{1}{4}}-1]e^{-\frac{1}{4}\kappa t})^4}}\ . \tag{4}$$

Using equation 4, the calculation of tumor growth can be made using only diameter.

The maximum diameter of a tumor X_{max} is $\lim_{t\to\infty} X(t) = 128$ mm.

2.4 The growth rate κ

In the equation for cancer mammae tumor growth (4), κ is the relative, and individual, growth rate. Following Weedon-Fekjær et. al. [1], it is assumed that κ varies between individuals according to the log normal distribution (5). Given the underlying parameters μ and σ^2 , pseudo random numbers from the log normal distribution can be generated using the function *rlnorm* in R. These parameters were estimated by Weedon-Fekjær et. al. [1] and are shown in Table 4.

Recall here that the log normal distribution has the following density function

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}\ , \tag{5}$$

which gives E[X] $= e^{\mu+\frac{\sigma^2}{2}}$ and Var[X] $= (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$.

Table 4: Estimated parameters of the log normal distribution for the growth rate κ, estimated by Weedon-Fekjær et. al.

	μ	σ
50 - 59	1.38	1.36
60 - 69	0.70	1.18
All ages	1.07	1.31

To visualize the effect of the different values of the parameters, Figure 10 shows tumor growth calculated by equation (4) for the expected value of κ in the considered age groups. As seen, there is a big difference in average tumor growth. The two lines in each color, indicate the average simulated κ and the average simulated κ after truncation, described in the next paragraph.

Table 5: Mean and standard deviation of the log normal distribution with parameters from table 4 and simulated values based on 1 000 000 simulations for the truncated distribution.

Expected	Mean ($e^{\mu+\frac{\sigma^2}{2}}$)	SD ($\sqrt{(e^{\sigma^2}-1)e^{2\mu+\sigma^2}}$)
50 - 59	10.02217	23.19711
60 - 69	4.039819	7.025683
All ages	6.876089	14.6878

Simulated	Mean	SD
50 - 59	8.538821	12.68382
60 - 69	3.974879	6.209643
All ages	6.337193	10.01364

The log normal distribution with its parameters estimated by Weedon-Fekjær et al. [1], has a heavy tail. Figure 11(a) shows 40 000 realizations of the log normal distribution and as seen, though most values are smaller then 100, some are very large. These large values of κ correspond to tumors that grow unrealistically fast. In Figure 11(b) growth of tumors with $\kappa = [100, 150, 200, 250]$ are shown. As an example, a tumor with $\kappa = 250$ (blue) will grow from size 0 mm to $X_{max} = 128$ mm in three months. This is not realistic.

To remove these unrealistic values, a truncated log normal distribution is used. Somewhat arbitrarily we chose to truncate at the value 100 for κ. In practice, each drawn value of κ is checked. If it is larger then 100, a new value is drawn until a value below 100 is obtained. Since these large values are rare, the mean after truncation is hardly changed. A simulation of 1 000 000 realizations of κ, showed that the mean changed from 6.876 to 6.337 after removing values of κ larger than 100.

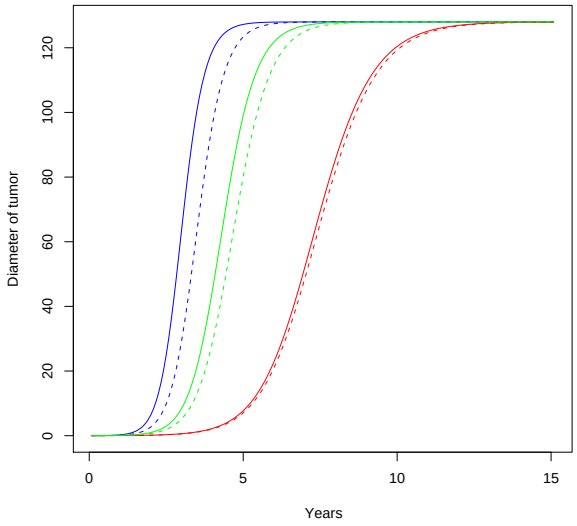
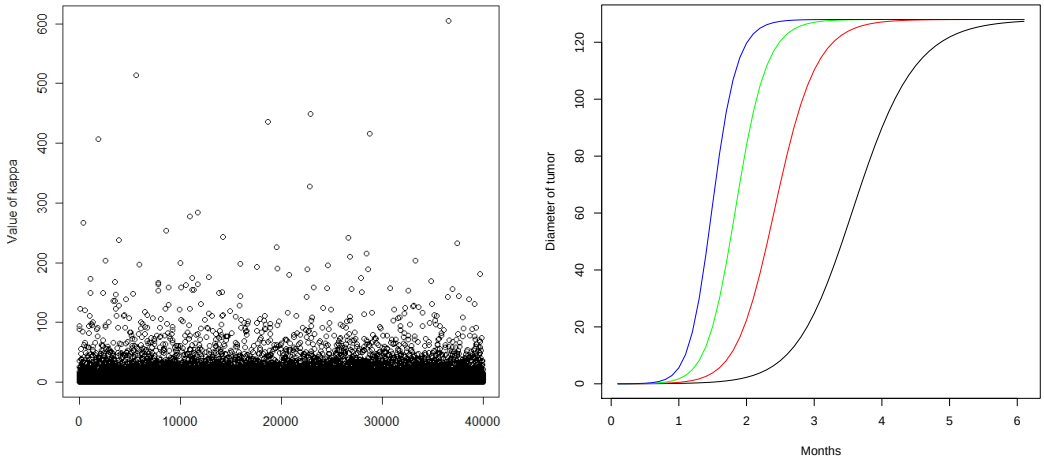


Figure 10: The tumor growth curves for the mean value of κ before and after truncation for all considered age groups. The solid lines show the non-truncated values, while the dotted lines show the growth curves after truncation. Blue: 50-59, red: 60-69 and green: all.



(a) (b)

Figure 11: a) 40 000 realizations of κ with $\mu = 1.07$ and $\sigma = 1.31$ and b) tumor growth curves for $\kappa = [100, 150, 200, 250]$

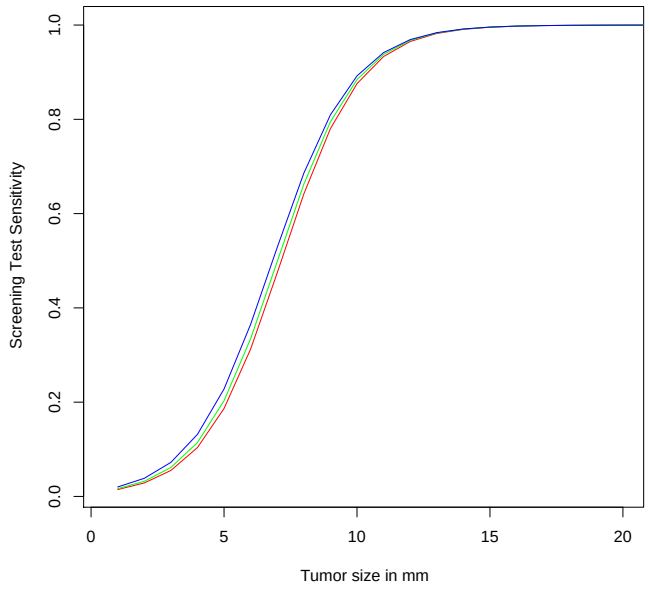


Figure 12: The screening test sensitivity function, $S(X)$. Blue: 50-59, red: 60-69 and green: all.

2.5 Sensitivity of screening tests

The screening test sensitivity (STS) is a measure of how likely it is to find a tumor of a given size at screening. Since larger tumors are easier to find on mammography screening, it is natural to use a function that increases with increasing tumor size. A logistic function is used by Weedon-Fekjær et. al. [1] to describe STS.

More preccisely, STS, the probability of finding a tumor with diameter X (in mm) at screening, is defined as

$$S(X) = \frac{exp\left(\frac{X-\beta_2}{\beta_1}\right)}{1+exp\left(\frac{X-\beta_2}{\beta_1}\right)}, \tag{6}$$

where β_1 defines how fast sensitivity increases by tumor size and β_2 relates STS to tumor size. β_1 and β_2 have been estimated using maximum likelihood by Weedon-Fekjær et. al. [1]. The estimated values of β_1 and β_2 for the different age groups are displayed in Table 6. The function is displayed in Figure 12 for the considered age groups. As seen, the STS function is very similar for the different age groups.

Table 6: Parameters of the screening test sensitivity

	β_1	β_2
All ages	1.47	6.51
50 - 59	1.50	6.33
60 - 69	1.46	6.65

2.6 Estimation of the parameters

In sections 2.4 and 2.5, parameters of the log normal distribution and the STS-function were given. These parameters were all estimated with the maximum likelihood method by Harald Weedon-Fekjær et al. [1]. The following model parameters, assumptions and data were used:

Assumptions

- Tumor growth curves follow a generalized logistic function
- Tumor growth κ varies between individuals following a log normal distribution
- No regressive tumors; once a tumor starts growing it does not stop
- Screening test sensitivity (STS) is a continuous increasing function of tumor size, following a logistic function
- Cancer incidence without screening, and size distribution of clinically detected tumors, can be estimated from historic data, with an added correction for incrased use of HRT use.

Model parameters

- μ Tumor growth parameter
- σ Variation in tumor growth rates (log normally distributed)
- β_1 Relating STS and tumor size
- β_2 Defining how fast STS increases with tumor size.

Data

- Data on tumor sizes found at screening were collected from the NBCSP (1995-2002)
- Observed cancer incidence and size distribution at screening and in the following years were collected from NBCSP and Statistics Norway (1995-2002)
- Assumed tumor sizes at diagnoses without screening are from Haukeland hospital, before the initiation of the screening program (1985-1994)
- The expected incidence of cancer mammae without screening was estimated from time-trends using the Norwegian Cancer Registry data (1978-1997).

The likelihood function

$L(\mu, \sigma, \beta_1, \beta_2) =$

$$\prod_{\text{all size groups } i} P(\text{observed number of cases at screening in different size groups } i \mid \mu, \sigma, \beta_1, \beta_2) \cdot$$

$$\prod_{\text{all observed intervals } j} P(\text{observed number of interval cancers } j \text{ months after screening} \mid \mu, \sigma, \beta_1, \beta_2)$$

2.7 Clinical detection diameter

Cancer mammae tumors are not only found by mammography screening. A lot of tumors are also found clinically. Clinical detection is mostly the result of a clinical symptom, i.e. feeling a bump, felt by the patient or a doctor. When running a simulation of cancer mammae development, a distribution of clinically detected tumors is needed. Since tumors removed are measured in diameter, it is natural to look for a distribution that could describe these sizes. Data on tumor size of clinically found tumors have been registered in Haukeland hospital, Norway, before the official screening program started (1985 - 1994). These data can therefore be used to estimate an underlying distribution of size of clinically found tumors. These data are illustrated (as a smoothed curve) in Figure 13.

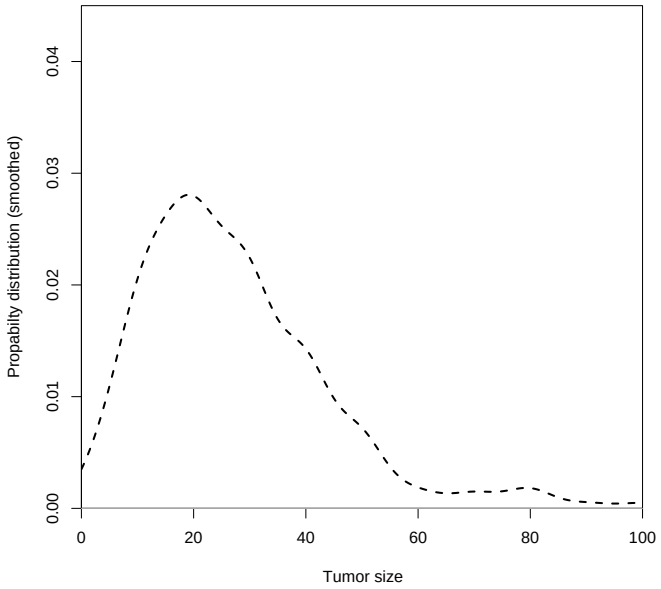


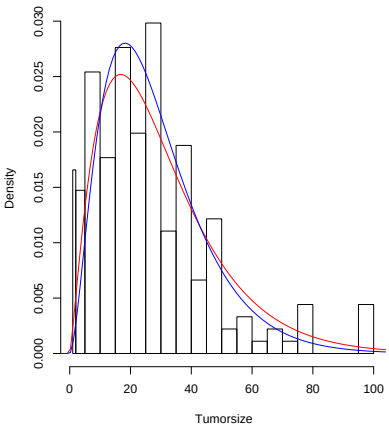
Figure 13: The tumor size distribution of tumors found clinically at Haukeland hospital before the official screening program started for the age group 'all'. Figure from Harald Weedon-Fekjær et al. [1]

Figure 14, shows histograms of the data sorted by age groups and using bars of length 5 mm corresponding to diameters 5, 10, 15 etc. It can be seen that measurements of tumors are not accurate. Especially for large tumors, there are unrealistically many observations in size groups 70, 80, 90 and 100, in comparison to groups 75, 85 and 95. This is likely to be a result of rounding to the nearest ten mm.

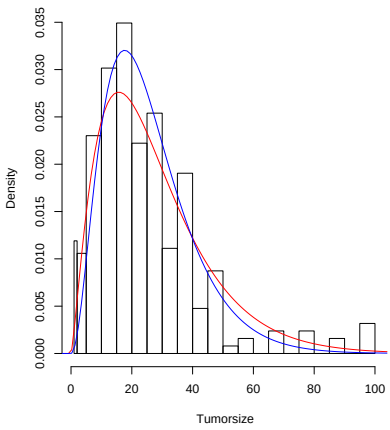
The data were fitted to gamma distributions (7), with parameters estimated using the method of moments described in section 3.1,

$$f(x;k,\theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{\frac{-x}{\theta}} \qquad \text{for x} > 0 \ , \tag{7}$$

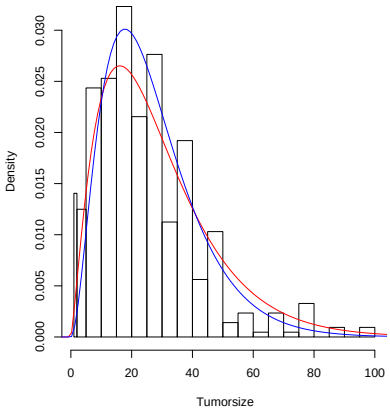
where k is the shape and θ is the scale parameter.



(a) 50 - 59



(b) 60 - 69



(c) All

Figure 14: The frequency of tumors found clinically in size groups of 5 mm and the gamma distribution with estimated parameters. The red line is the gamma distribution with the unadjusted parameters, while the blue line indicates the adjusted results.

Table 7: Parameters of the Gamma distribution

	Shape (k)	Scale (θ)	Shape (adjusted)	Scale (adjusted)
All ages	2.3	12.386	2.958	9.076
50 - 59	2.276	13.145	2.787	10.168
60 - 69	2.337	11.751	3.165	8.145

Table 7 shows the result of the parameter estimation. Two different sets of parameters are considered; a shape/scale and an adjusted shape/scale parameter. The adjusted parameters are based on data without using the values larger than 85 mm, while the unadjusted ones are based on all available data. Looking at Figure 14 again, two smoothed lines are plotted for each age group. The red line is the gamma distribution with the unadjusted parameters, while the blue line shows the adjusted results. The unadjusted density (red) has a heavy tail to pick up the large detection values, but fails to sufficiently pick up the "normal" values. Tumors larger than 85 mm are ignored in the simulation because such tumors often are results of denial of the patient. The adjusted parameters are used in the following.

3 Statistical methods

3.1 Method of moments

The method of moments, is a simple way of estimating parameters such as mean, variance or median from a known distribution. For the gamma distribution in equation 7, this is done by equating the known expected moments, $E[X] = k\theta$ and $Var[X] = k\theta^2$, to the observed mean and variance and solving the equations.

$$k\theta = \bar{X} \Rightarrow \hat{k} = \frac{\bar{X}}{\hat{\theta}}$$

$$k\theta^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2 \Rightarrow \hat{\theta} = \frac{\frac{1}{n-1}\sum_i (X_i - \bar{X})^2}{\bar{X}}$$

3.2 Poisson process

The Poisson process is a stochastic process, counting the number of events in continuous time, for example telephone calls received by an office, the number of typing errors or the arrival times of customers at a service center. In this process events must occur according to the following properties:

The Poisson process has no memory; the number of outcomes occurring in one time interval, or other unit of measurement used, is independent of the number of events occurring in any other disjoint time interval.

The probability that a single outcome will occur during a very short time interval, or other unit of measurement used, is proportional to the length of the time interval and does not depend on the number of outcomes occurring outside this time interval.

The probability that more than one outcome will occur in a short time interval, or other unit of measurement used, is negligible.

It is reasonable to assume that the initiation of cancer mammae in a population of women follows a Poisson process.

The Poisson process is characterized by its rate parameter λ, which is the expected number of events occurring per unit time. In 2006, there were 1064 new cancer cases among the 520 196 Norwegian women aged 50 - 69. The simulation uses months as time steps, which means approximately 89 new cases per month, among the 520 196 women. That means that the rate parameter $\lambda = n \cdot \frac{89}{520196}$, where n is the number of women in the simulation.

Part II
Simulation

Introduction

The goal of this part, is to describe steps needed to make a realistic simulation of a Norwegian female population.

4 R

R is a free, statistical software package for computing and graphics. It is a GNU project which is similar to the S language and environment. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered in R. Many important statistical methods are pre-programmed in R. Some examples are linear and nonlinear modeling, classical statistical tests, time-series analysis, classification and clustering. The pseudo random number generator from different distributions is particularly important in the simulation of cancer mammae.

All simulations and plots in this paper are obtained using R.

5 How the simulation works

When the simulation is initiated, the population consists of 25 000 women, who are all cancer free. A separate simulation is run for women in age groups 50-59, 60-69 and both (50-69). These specific age groups are chosen because the estimated parameters are based on data from these groups.

As the simulation is initiated, for each woman a latent growth rate κ and a clinical detection diameter is drawn from the distributions given in section 2.4 and 2.7. Knowing the growth rate κ and at what size the tumor will be found clinically, equation 4 gives the full path of the tumor growth.

Since it is not realistic that all women in the population get cancer at the same time, or even get cancer at all, the event that a women gets cancer is assumed to follow a Poisson process. With the assumptions of the Poisson process, it follows that new cases of cancer mammae are randomly distributed throughout the year. At each time step (one month) in the simulation, a random selection of women get cancer. This is drawn by the pseudo random number generator of the Poisson distribution, *rpois* in R.

For the three considered age groups, different screening intervals are tested. These intervals are:

- Screening every year
- Screening every other year
- Screening every three years
- Screening every five years
- Screening every ten years

When screening takes place, all women are entered and the probability that their tumor (if one is present) is detected (STS) is calculated. This is described in more detail in section 5.2.

5.1 Reaching a stationary distribution

Before a realistic simulation of screening can be made, the population needs to reach a stationary state, meaning that the distribution of tumor sizes is approximately the same at all time steps. To reach this state, the simulation must go through a burn-in period.

As shown in section 2.4, some simulated values of κ may become very large and are therefore ignored (truncated). On the other hand, a lot of values of κ are very small. About 2% of the values are less than 0.2. Figure 15 shows tumors with small values of κ growing for 100 years. As seen, a tumor with $\kappa = 0.1$ or $\kappa = 0.2$ is less then 10 mm in diameter after 100 years.

These slow growing tumors are hence a problem when trying to reach a stationary state. At every time step in the simulation, new tumors may be initiated. When a very slow growing tumor is initiated, it will stay in the simulation for a long time without being detected neither clinically nor at screening, because of its small size. After many time steps, this will result in a large number of small, very slow growing tumors. This prevents the simulation from reaching a stationary state as the number of small tumors increases. The consideration of such tumors is also unrealistic, since we are looking at women aged 50 years or older, and the expected lifespan of a (Norwegian) woman is 81,5 years according to Statistisk sentralbyrå in 2008.

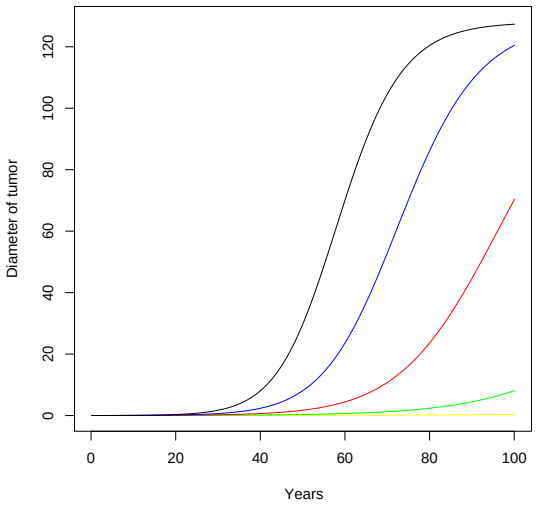


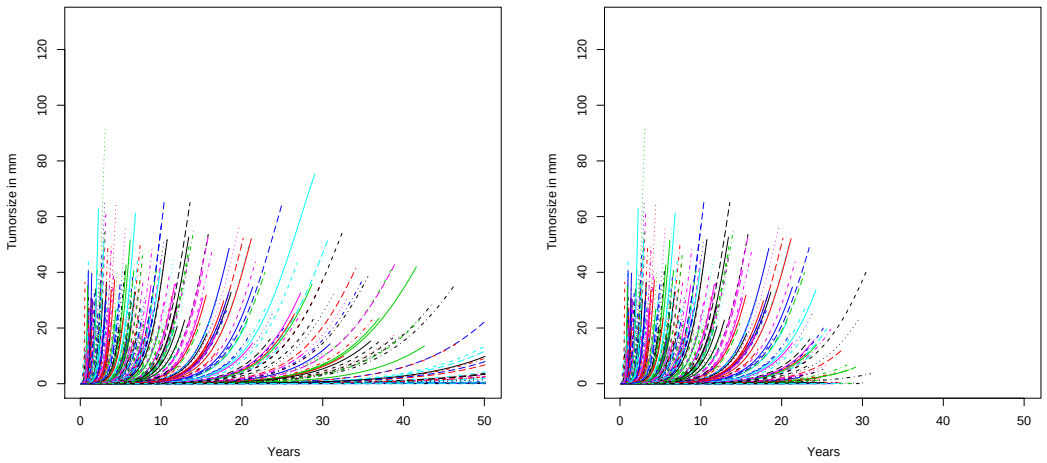
Figure 15: Tumor growth curves for tumors with $\kappa = [0.1, 0.2, 0.3, 0.4, 0.5]$

This problem can not be solved with truncation as for the large values of κ, since there are many slow growing tumors in the population. The solution to the problem was to introduce a time limit. With a time limit, it is certain that no tumor will keep growing unrealistically long. This is illustrated in Figure 16. Figure 16(a) shows all tumors growing without a limit from time 0 and 16(b) shows the same tumors after introducing a time limit.

The time limit was set to be normally distributed with parameters stated in Table 8. In practice, all tumors are given a time limit. If the tumor is not detected clinically nor at screening, the tumor is removed from the simulation.

Table 8: Parameters of the normal distribution giving a time limit to tumors. The time unit is months.

	Mean (μ)	SD (σ)
50 - 59	300	30
60 - 69	180	30
All ages	240	30



(a) Without a growth limit

(b) With a growth limit

Figure 16: The plot shows 400 randomly selected tumor growth curves all starting at time 0. Each colored line represents the tumors growth curve for one woman. The height indicates the size of the tumor and the length along the x-axis, describes how long it takes before the tumor is found clinically. Age group 'all'.

5.2 Screening

When the population has reached a stationary state, screening can take place in the simulation. Screening occurs in the same month every year and the entire population is screened in the same time step. Berit Damtjernhaug at the Cancer Registry of Norway, found the true turnout on mammography screening to be 76.2% in Norway, based on data from NBCSP (1996 - 2005) [14].

At screening, every woman has a mammography. The STS function, as given in equation (6), is used to determine how likely it is that a tumor is detected. Since a tumor is either found or not, the binomial distribution by the function *rbinom* in R, with parameters $n = 1$ and $p = STS$, is used to decide whether a tumor is found or not, after STS has been calculated.

At the lower end of the scale, for diameters close to 0, the STS function in equation (6) does not work well. This is illustrated in Figure 12. If a tumor with diameter 0, thus no tumor, is entered to equation 6, the probability is 0.0118 that it found on mammography screening. Since most women do not have a tumor, this results in a lot of positive tests when there is no tumor to be detected.

To mend this problem, the STS for all tumors less then 0.5 mm in diameter is set to 0. This is reasonable since it is not likely that any tumor less than 0.5 in diameter will be found at screening.

5.3 Saved time because of screening

Looking back to Figure 5, the idea behind mammography screening is to detect cancer mammae tumors earlier than they would have been if not screened. An interesting result is therefore to see how much earlier, if earlier at all, the tumors are actually found if screened in the simulation.

To calculate how much time is saved with different screening intervals, different simulations had to be run. For screening every year, all women were screened for the first time at $t = 12$ (months) after burn-in. Whether a tumor is detected or not, is determined as described in section 5.2.

If a tumor is detected at screening, the time until detection is registered. This time is then compared to the time it would have taken before the tumor would have been found clinically, which can be computed using the gamma distributions described in section 2.7 and the given value of κ. The difference in times is recorded as the saved time for that woman.

In the case where there is no tumor, or the tumor is too small to be detected at screening, the simulation moves on to the next screening and the process is repeated. Large tumors will be detected in the interval between the screenings if it reaches its clinical detection diameter in the interval.

Part III
Results

6 Results from simulation

6.1 The size distribution of tumors

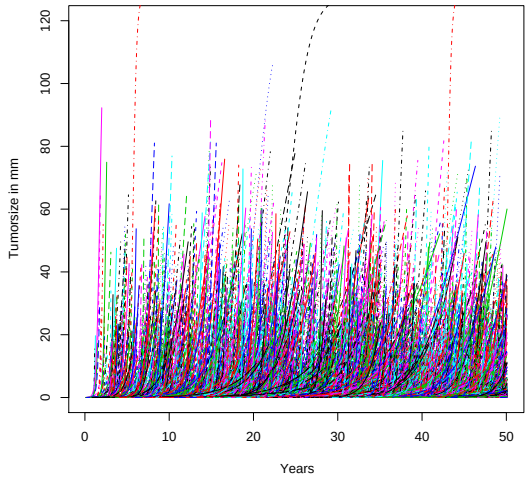
In section 5.1 the necessity for reaching a stationary state was discussed. Figures 17(a), 17(b) and 17(c) show tumor growth after the initiation by the Poisson process. It looks as if the population is stable after 10 years.

The size distribution of tumors, after reaching a stationary state, at nine given times in the age group 'all', are plotted in Figure 18. The times were randomly selected, for $t > 15$ years (180 months), as the population then seems to be in a stationary state. The plot only shows tumor sizes between 0 and 10 mm, because the number of larger tumors is very low. The plots show similar curves, and it is therefore reasonable to assume that a stationary state has been reached and that the values originate from the same distribution. Other age groups show similar results.

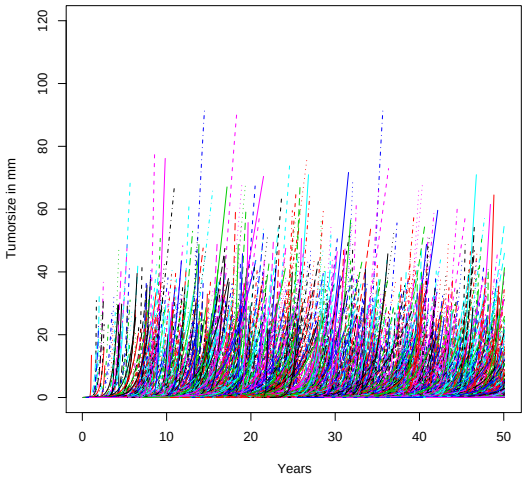
The tumor sizes were fitted to a gamma distribution (7). As in section 2.7, parameters were estimated by the method of moments based on data from the nine selected times. Table 9 shows the estimated mean and variance for all age groups and the estimated parameter k and θ.

Table 9: Mean tumor size and variance (in mm) for all age groups and estimated parameters of the gamma distribution.

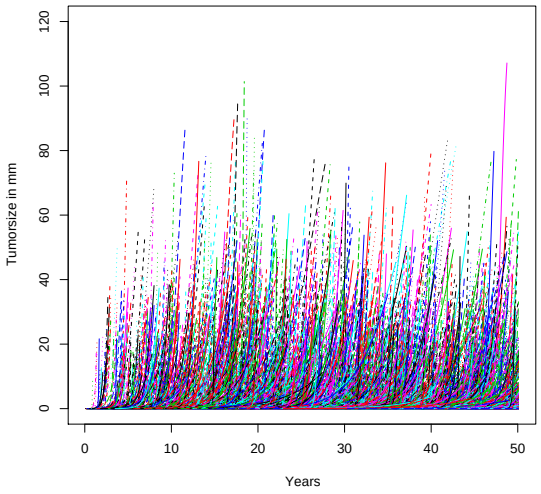
	Mean ($k\theta$)	SD ($\sqrt{k\theta^2}$)
50 - 59	3.035	7.641
60 - 69	1.831	5.268
All ages	2.449	6.469
	k	θ
50 - 59	0.158	19.234
60 - 69	0.121	15.156
All ages	0.143	17.093



(a) 50-59



(b) 60-69



(c) All

Figure 17: For the three considered age groups, growth curves initiated according to the Poisson process are shown. The plots show 2000 randomly selected tumors. Each colored line represents the tumor growth curve for one woman. The height indicates the size of the tumor and the length along the x-axis, describes the time it takes before the tumor is found clinically.

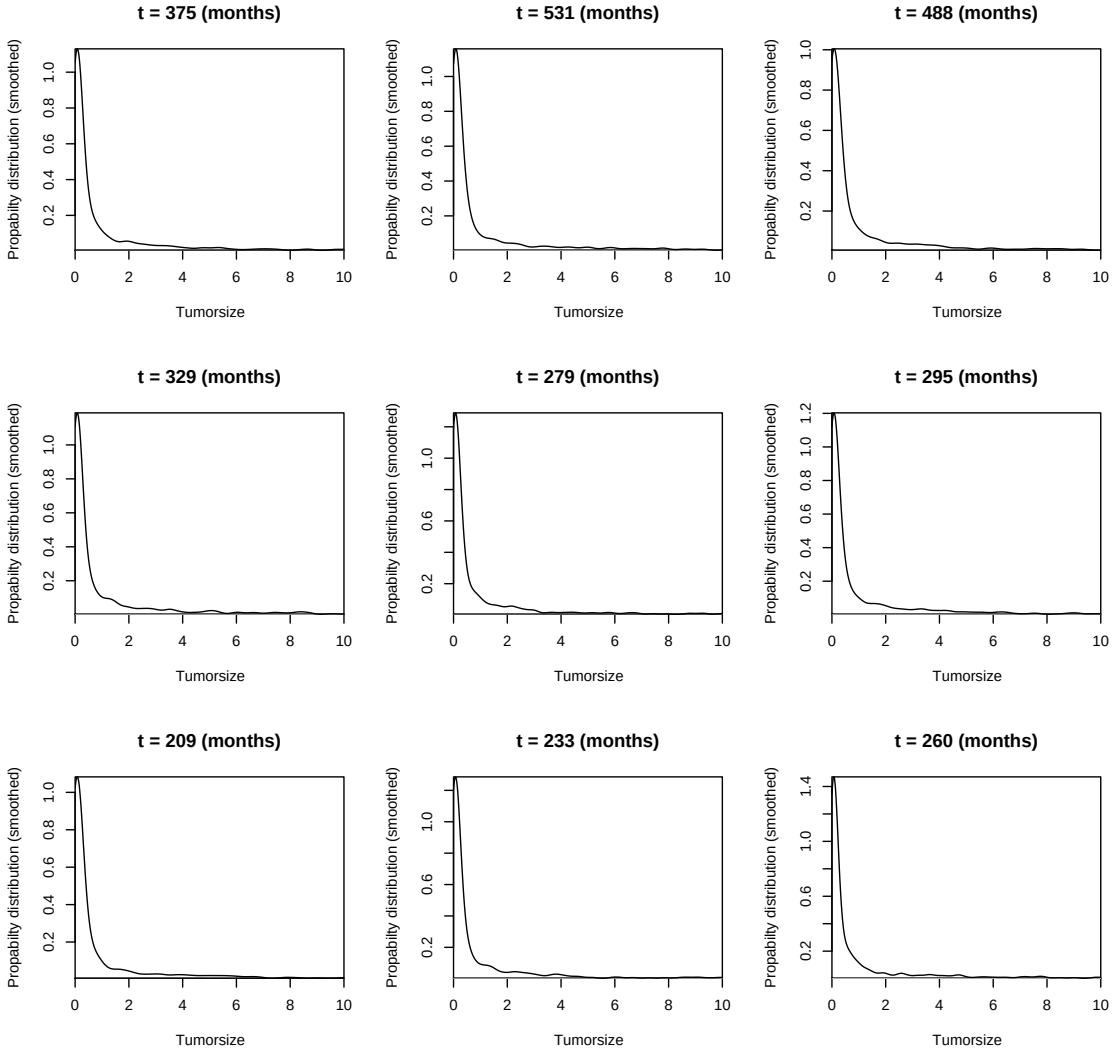


Figure 18: The size distribution of tumors with diameter > 0 at random times during the simulation.

When screening is introduced to the population, it is to be expected that the tumor size distribution is changed. It is also reasonable to assume that the size distribution is different when screening is performed every year, compared to screening at longer intervals. It is also of interest to find the difference in size distributions before and after screening.

A simulation to determine these distributions was made. Tables 10 and 11 show the results. The gamma distribution was found to give a very good fit also for this size distribution.

Table 10 shows the estimated mean, variance and parameters one time step `before` screening. The distribution before the first screening is assumed to follow the distribution of a stationary population without screening and these data are therefore not included here. For all consecutive screenings (2nd, 3rd, 4th, and so on), the distribution is assumed to be the same. The calculation is made in the time step (one month) before screening takes place in the simulation.

As expected, the mean tumor size before screening becomes larger with infrequent screening. This is reasonable, because tumors have more time to grow between each screening. Comparing the last lines in Table 10, it can be seen that for screening every five or every ten years, the mean tumor size is almost as large as in a population without screening.

It is interesting to see that the parameter k, the shape parameter, is very similar for all the considered screening frequencies. The Coefficient of variation, CV, in the gamma distribution is $\frac{1}{k} = (cv)^2$. Since k is very similar for all screening frequencies, within each age group, the CV is also similar.

Table 11 shows the same values `after` screening. It is interesting to see that the tumor size distribution after screening is very stable in different screening frequencies. This is a result of how mammography screening works. Since a tumor can first be seen on a mammography x-ray after reaching a certain size, the size distribution of tumors found at screening will be very similar, independent of the screening frequency.

The difference between the age groups can also be seen from the tables. Tumors in younger women (50 - 59) are larger, on average, and have a higher variance than in older women, especially before screening.

Table 10: Mean tumor size and SD (in mm) one time step (one month) **before** screening for all age groups and screening intervals, and estimated parameters of the gamma distribution.

	Age group		
	50-59	60-69	All
Screening frequency	Mean - SD	Mean - SD	Mean - SD
Every year	1.293 - 3.167	0.953 - 2.613	1.187 - 3.010
Every two years	1.919 - 5.032	1.451 - 4.039	1.682 - 4.343
Every three years	2.309 - 5.794	1.629 - 4.463	2.060 - 5.199
Every five years	2.629 - 6.426	1.871 - 5.265	2.322 - 5.878
Every ten years	2.964 - 7.581	1.900 - 5.201	2.755 - 7.078
Never	3.035 - 7.641	1.831 - 5.268	2.449 - 6.469

	Age group		
	50-59	60-69	All
Screening frequency	k - θ	k - θ	k - θ
Every year	0.166 - 7.759	0.133 - 7.165	0.155 - 7.631
Every two years	0.145 - 13.186	0.129 - 11.245	0.150 - 11.211
Every three years	0.159 - 14.543	0.133 - 12.224	0.157 - 13.116
Every five years	0.167 - 15.702	0.126 - 14.814	0.156 - 14.877
Every ten years	0.152 - 19.391	0.133 - 14.232	0.151 - 18.181
Never	0.158 - 19.234	0.121 - 15.156	0.143 - 17.093

Table 11: Mean tumor size and variance (in mm) one time step (one month) **after** screening for all age groups and screening intervals, and estimated parameters of the gamma distribution.

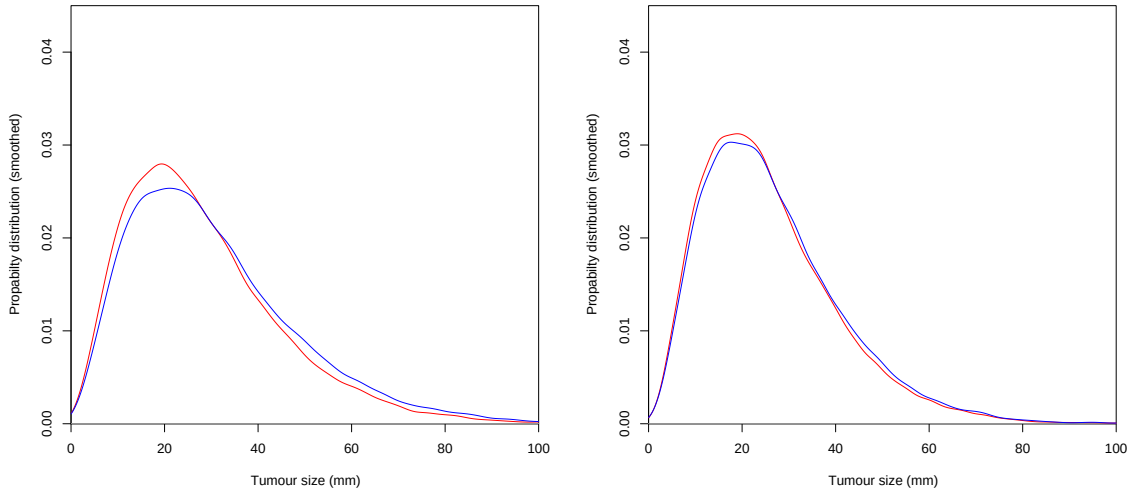
	Age group		
	50-59	60-69	All
Screening frequency	Mean - SD	Mean - SD	Mean - SD
Every year	0.784 - 1.628	0.605 - 1.386	0.732 - 1.593
Every two years	0.839 - 1.745	0.662 - 1.522	0.758 - 1.570
Every three years	0.868 - 1.774	0.659 - 1.517	0.788 - 1.685
Every five years	0.876 - 1.775	0.667 - 1.516	0.781 - 1.679
Every ten years	0.828 - 1.823	0.664 - 1.526	0.748 - 1.558
	Age group		
	50-59	60-69	All
Screening frequency	k - θ	k - θ	k - θ
Every year	0.232 - 3.381	0.190 - 3.179	0.211 - 3.464
Every two years	0.231 - 3.626	0.189 - 3.500	0.233 - 3.250
Every three years	0.239 - 3.624	0.189 - 3.487	0.219 - 3.599
Every five years	0.244 - 3.597	0.193 - 3.446	0.216 - 3.607
Every ten years	0.206 - 4.014	0.189 - 3.506	0.230 - 3.242

6.2 Clinical detection

In section 2.7, parameters of the gamma distribution were estimated from real data to be used in simulation of a clinical detection size. Figures 19(a), 19(b) and 19(c) show the estimated gamma distribution of size of clinically detected tumors (red) and the tumor size distribution found in the simulation (blue), for all age groups. To make comparison easier, Figure 13 from section 2.7 is repeated as Figure 19(d).

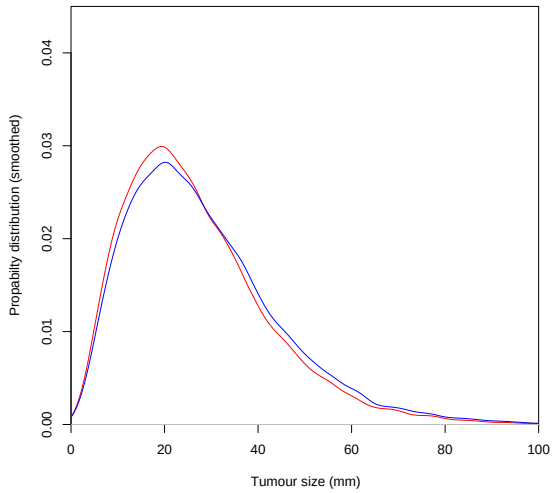
Figure 19(d) shows the non-parametrically estimated tumor size distribution of tumors found clinically before the official screening program started. As can be seen, the estimated gamma distribution (red) and the tumor size distribution found in the simulation (blue) are very similar. The tumor size distribution found in the simulation is also very similar to the size distribution estimated from the observed data. This is reasonable, since the gamma distribution is estimated from the Haukeland data and the simulated values are drawn from this gamma distribution.

In the simulation, a tumor is detected clinically once its size is larger or equal to the detection size drawn from the gamma distribution. This discretization of the time steps causes the slight bias seen in Figure 19.

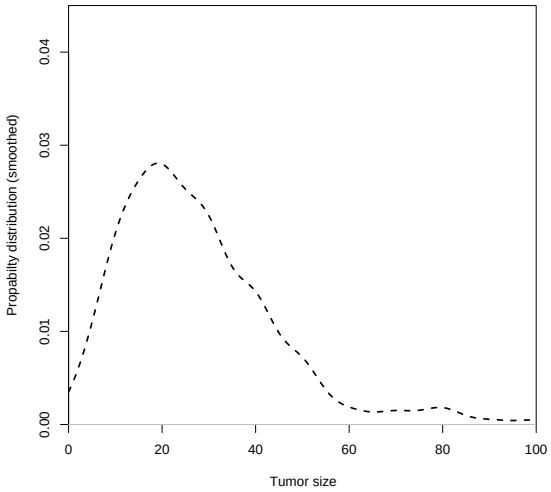


(a) 50 - 59

(b) 60 - 69



(c) All



(d) All, form Haukeland hospital.

Figure 19: The estimated gamma distribution of size of clinically detected tumors in red compared to the size distribution found in the simulation in blue.

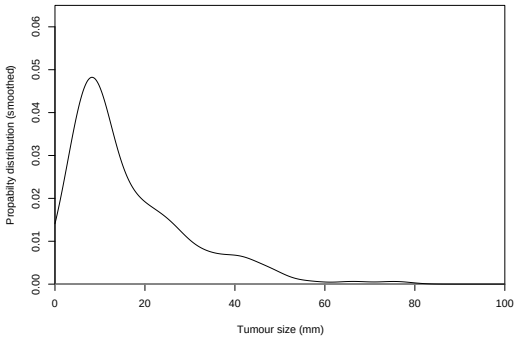
6.3 Detection at screening

In section 1.7 screening by mammography was described. Figure 6 showed the size distribution of tumors found at screening, based on data from the NBCSP (collected 1995-2002). The figure is repeated as Figure 20(d). These data were strictly from the first appearance at screening of women aged 50-69 reporting no earlier screening history.

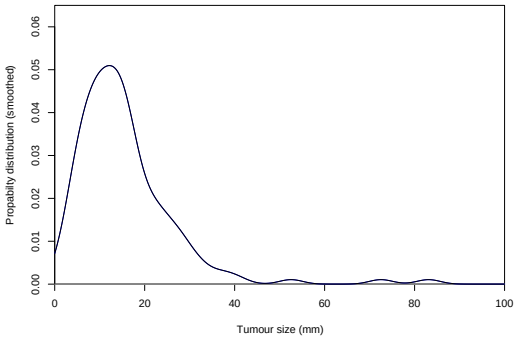
When running the simulation, the "women" entering the first screening have the same properties; it is their first appearance at screening and they have not been screened before.

The size distribution of tumors found at screening is very dependent on the STS-function, since this function decides which tumors are detected at screening.

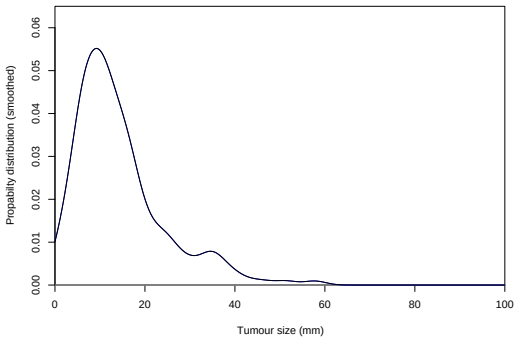
Comparing Figure 20(c) and 20(d), it can be seen that the two densities are similar. Unlike the clinical detection size, which is estimated from observed cancer mammae data, the detection size at screening is just observed in the simulation. Seeing that the size distribution in the simulation is similar to the observed data from the NBCSP, is therefore a good indication that the simulation, and especially the STS-function, is working properly.



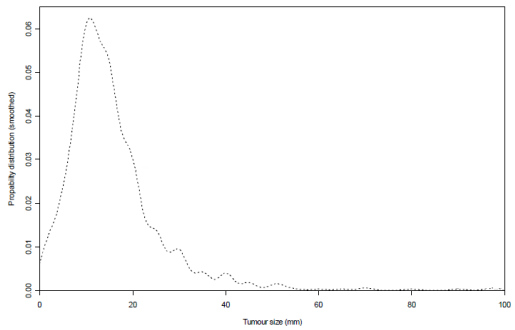
(a) 50 - 59



(b) 60 - 69



(c) All



(d) All, from NBCSP.

Figure 20: Size distribution of tumors found at first screening. (a), (b) and (c) are from the simulation, while (d) is from the observed data (NBCSP)

6.4 Reduction in number of cancer cases

Table 12 shows the number of cancer cases found at various screenings in the simulation.

As seen, the number of tumors found at the first screening is similar within each age group. This is as expected as a stationary state is reached before the first screening. The table also shows that more tumors are detected at screening in younger women in the first screening and also in the following rounds of screening. This is natural as tumors in younger women grow faster, on average.

With screening every year, the number of tumors found at screening stabilizes on a lower level than with more infrequent screening. This is in accordance with the results found in section 6.1.

Figures 21 and 22 show tumor growth curves with and without screening for different screening intervals. The grey lines indicate when screening is initiated and when it ends, respectively. How many screenings that take place between the to grey lines, is determined by the screening frequencies. For practical purposes, the plot for screening interval 'every five years' is left out.

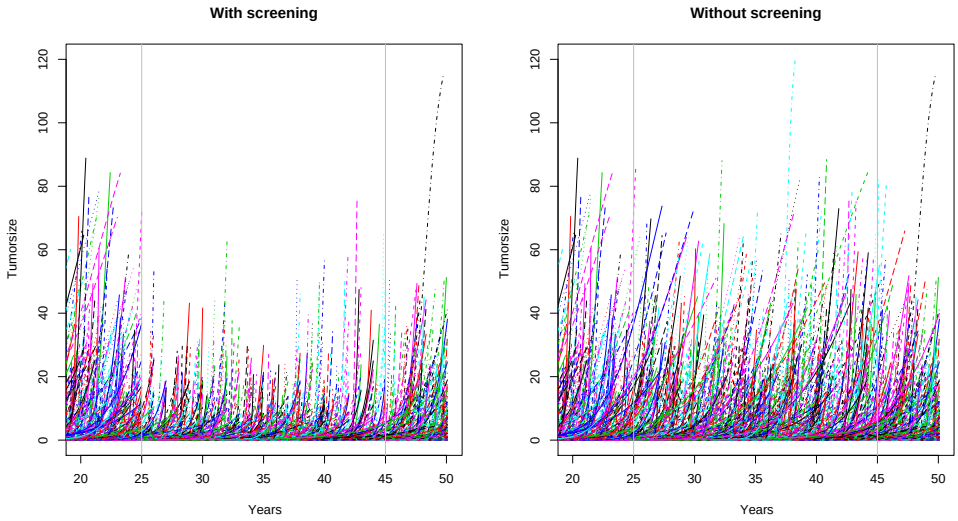
Figure 21(a) shows the situation for screening every year. As can be seen, a lot of tumors are removed at the first screening; the colored lines are removed at the first grey line. Also, in the time interval where screening is active, between the two grey lines, tumors are smaller (the colored lines are lower). This indicates that tumors are found smaller (and thus earlier) than they would have been if not screened. For screening every two years, shown in 21(b), the results are similar, but the reduction in tumor size is less.

When screening is less frequent, the difference between the growth curves with and without screening become more alike because tumors have more time to grow between each screening.

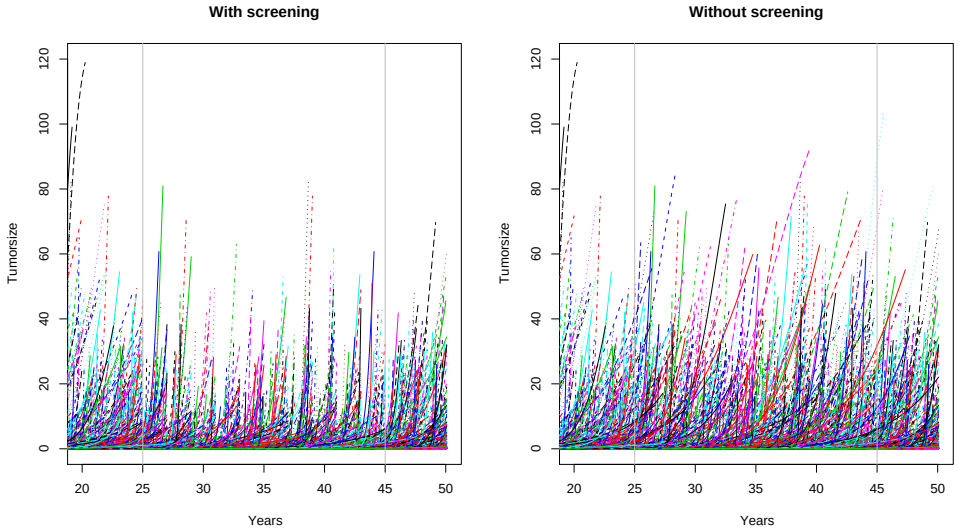
Another interesting point, is to see how fast after the last screening the tumor growth curves are back on the same level as when no screening was performed.

Table 12: Number of cancer cases found at the first six screenings for the considered screening frequencies and age groups

Age group	50-59					
Screening freq.	1. screen	2. screen	3. screen	4. screen	5. screen	6. screen
Every year	1233	699	664	688	617	648
Every two years	1200	928	923	881	832	759
Every three years	1207	1033	1044	983	916	824
Every five years	1220	1146	1088	1054	904	-
Every ten years	1202	1148	1142	-	-	-
Age group	60-69					
Screening freq.	1. screen	2. screen	3. screen	4. screen	5. screen	6. screen
Every year	842	566	512	503	481	487
Every two years	846	720	690	673	687	602
Every three years	831	825	726	741	771	655
Every five years	866	799	824	730	646	-
Every ten years	825	792	792	-	-	-
Age group	All					
Screening freq.	1. screen	2. screen	3. screen	4. screen	5. screen	6. screen
Every year	1049	619	635	550	614	630
Every two years	1026	874	803	820	801	741
Every three years	1018	909	973	882	884	763
Every five years	1089	1022	974	885	854	-
Every ten years	1058	1022	973	-	-	-

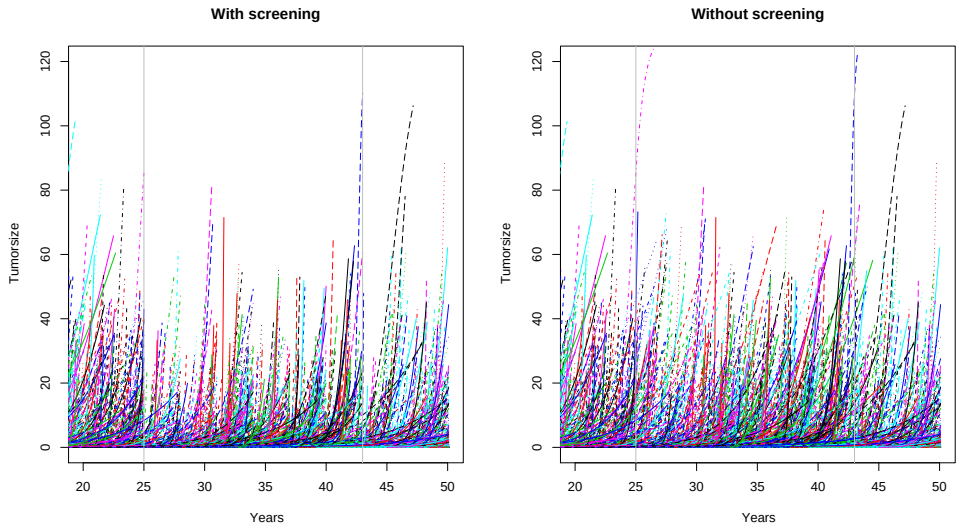


(a) Screening every year

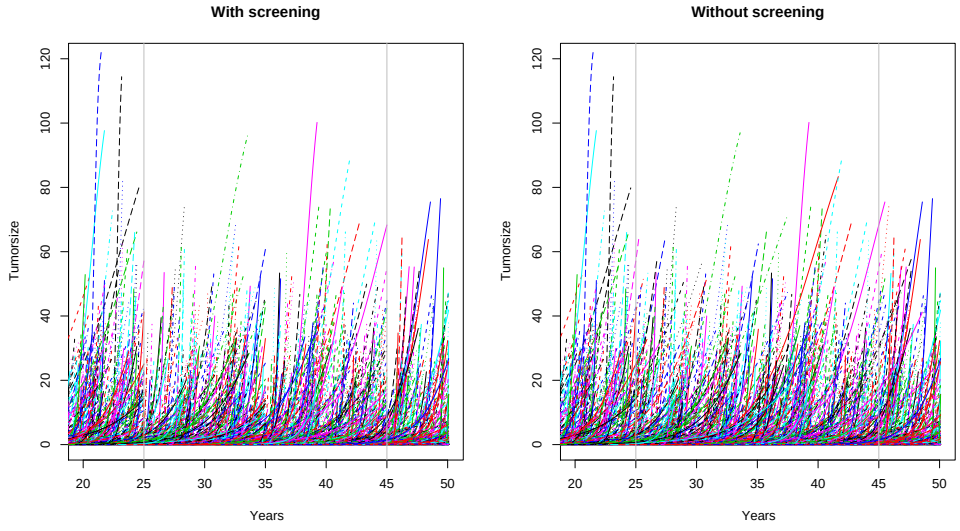


(b) Screening every two years

Figure 21: Tumor growth curves with or without screening. Each colored line represents the tumor growth curve for one woman. The height indicates the size of the tumor and the length along the x-axis, describes the time it takes before the tumor is found clinically. The grey, vertical lines indicate when screening is initiated and when it ends, respectively. Age group 'all'.



(a) Screening every three years



(b) Screening every ten years

Figure 22: Tumor growth curves with or without screening. Each colored line represents the tumor growth curve for one woman. The height indicates the size of the tumor and the length along the x-axis, describes the time it takes before the tumor is found clinically. The grey, vertical lines indicate when screening is initiated and when it ends, respectively. Age group 'all'.

6.5 Values of the growth rate κ

Table 13 shows the average value of the growth rate κ for tumors found clinically and at screening for all age groups and screening frequencies.

The results show that tumors found clinically mostly have larger values of κ than tumors found at screening. An exception can be seen in the age group '60 - 69', where values of κ are larger in tumors found at screening than in tumors found clinically for screening frequencies every and every two years. This is realistic because an aggressive tumor (large κ) has a shorter MST and is more likely to be found clinically before screening takes place.

The table also shows that more infrequent screening results in somewhat less aggressive tumors being found on screening. If screening takes place every ten years, even less aggressive tumors will have time to become so large that they are found clinically and only slower growing tumors will be detected at screening.

The difference between the age groups, can be explained by the fact that the average κ in the age groups is very different. Table 5 showed the mean of κ and it can be seen that the mean is similar to the values of detected tumors.

Table 13: Average growth rate κ for tumors found clinically and at screening

	Age group		
	50-59	60-69	All
Screening frequency	Clinical - Screening	Clinical - Screening	Clinical - Screening
Every year	10.6389 - 6.0522	3.6721 - 4.4488	7.1077 - 5.3299
Every two years	10.3601 - 5.2514	3.9041 - 4.0512	7.1899 - 4.5916
Every three years	9.7859 - 4.9654	3.9779 - 3.9054	7.0135 - 4.3325
Every five years	9.5258 - 4.6966	3.9708 - 3.8498	6.8169 - 4.2813
Every ten years	9.1486 - 4.4453	3.9713 - 3.8992	6.6280 - 4.1179

6.6 Saved time because of screening

In section 5.3 the method of recording saved time was presented. In Tables 14 and 15, two different results are given.

Table 14 shows measurements of time saved for tumors that were actually found at screening. Thus, tumors that are found clinically in the intervals between screenings are not considered. These results are interesting, and give a realistic view, because a woman who has found a tumor clinically (in an interval) is not likely to participate in the normal screening program.

Table 15 shows measurements of time saved for all tumors and the standard deviation of these measurements.

As both tables show, the most time is saved if all women are screened every year and less time is saved with more infrequent screening. This is as expected. There is a slight difference between the age groups. Somewhat more time is saved by screening older women, which is likely to be a result of tumors growing slower in this age group.

The mean sojurn time (MST), described in section 1.7, tells us how long, on average, a tumor is of screening detectable size, before it is found clinically. The MST is estimated to 34.8 months (2.9 years) for the age group 'all' by Harald Weedon-Fekjær et al. [1].

The numbers in Table 14 are based on measurements of the time until a tumor would be found clinically, minus the time until the tumor is found on screening, given that it is not found clinically before screening. Table 14 shows that a tumor found at screening would be found clinically, on average, 19 months later, with screening every year in the age group 'all'. This is a reasonable result, because it implies that tumors, on average, are found approximately in the middle of the MST-interval.

The numbers in Table 15 are based measurements of the time until a tumor is found clinically, minus the time until the tumor would be found on screening. If the tumor is found later at screening than it would have been clinically, the result is a negative number and no time is saved by screening. Table 15 shows that a tumor found at screening would be found clinically, on average, 15 months later, with screening every year in the age group 'all'. The standard deviations are, though, fairly high.

For screening every year, it can be seen from the tables that the most time can be saved because of screening in all age groups. This is reasonable, because frequent screening allows less tumors to grow to clinically detection size. When screening is reduced to every second year, a lot less time is saved because of screening. The tumors that are actually found, are found about ten months earlier than they would have been without screening, while on average of all tumors only about 5 months is saved (dependent on the age group).

Screening every three years allows tumors to be found about eight months earlier if found at screening, but on average, tumors are found just as early clinically (the time until it would have been found clinically, minus the time until a tumor is found on screening is about zero).

If screening is only every five or every ten years, some time will still be saved if a tumor is found at screening, but the tumors will, on average, be found earlier clinically (the time until it would have been found clinically, minus the time until a tumor is found on screening is, on average, less than zero).

In general, the tables show that for tumors actually found on screening, some time is saved compared to no screening (when all tumors are found clinically). But, with screening less frequent than every two years, tumors are found as early or earlier without screening.

Table 14: Average saved time (in months) because of screening, with different screening intervals, for tumors being found on screening before being found clinically.

	Age group		
Screening frequency	50 - 59	60 - 69	All
Every year	19.9149	20.1934	19.0896
Every two years	10.0587	9.8625	9.4820
Every three years	8.81391	8.5083	8.3353
Every five years	6.64001	6.2254	6.1426
Every ten years	3.67423	3.6349	3.4334

Table 15: Average saved time (in months) and standard deviation because of screening, with different screening intervals. Negative numbers indicate that tumors will be found clinically earlier than at screening.

	Age group		
Screening frequency	50 - 59	60 - 69	All
	Mean - SD	Mean - SD	Mean - Sd
Every year	15.873 - 33.873	16.688 - 29.730	15.022 - 29.105
Every two years	4.367 - 20.874	6.046 - 19.902	4.836 - 19.725
Every three years	0.477 - 20.319	3.127 - 19.103	1.401 - 19.505
Every five years	-6.771 - 23.268	-3.276 - 21.254	-5.574 - 22.132
Every ten years	-23.577 - 36.820	-17.356 - 34.198	-21.475 - 35.759

Part IV
Discussion and conclusion

This paper takes advantage of present technology, both hardware and software, to simulate a large, synthetic population where issues normally not available for research because of practical limitations or ethical concerns, can be tested. The results from the simulation were compared to observed data on cancer mammae, and were found to be realistic and expected, verifying the underlying model and parameters estimated by Harald Weedon-Fekjær [1].

There are some aspects not covered in this paper. The results given are only an indication of which tumors are seen on a mammography x-ray; no further diagnosis is given. There is also no connection between how much time is saved because of screening and mortality rates. A measure of such would be a large improvement of the simulation study.

In the model, the individual growth rate κ and the gamma distributed clinical detection size are assumed to be independent. It would be interesting to find out if this assumption holds. My conjecture, is that an aggressive tumor (large κ) is found earlier clinically, and thus smaller, than a less aggressive (small κ) tumor.

A possible weakness is the assumption that a cancer mammae tumor is ball shaped. Tumors do not have perfect ball shapes. Also, the diameter of a tumor is roughly measured (as seen in the data from Haukeland hospital) and this might cause bias to the estimated parameters, and therefore to the calculations in this paper.

Several types of size distributions of tumors were found to be well fit with a gamma distribution. The size distribution of tumors found at screening in the simulation was similar to observed data from NBCSP. Because this distribution is based on the STS-function, its similarity with real life observations is a strong indication that both the simulation and the STS-function work well.

As expected, the size distribution of tumors found clinically in the simulation followed the same distribution as the observed data from Haukeland hospital. Also, the simulation found a strong reduction in number of cancer cases with frequent screening (every- or every second year), while less frequent screening results in a lower reduction.

The distribution of the growth rate κ varies between the age groups. It was found that tumors detected on screening, on average, are less aggressive (smaller value of κ) than tumors found clinically, except for screening every- and every two years in the age group '60 -69'.

An important question is how much time can be saved, on average, because of screening with different screening intervals. Two aspects were given; one only considering the tumors actually found at screening while the other included both tumors found clinically and at screening. The results were somewhat mixed. With screening every year, for women aged 50 - 69 (age group 'all'), tumors are found 15 months earlier than without screening, on average, with a standard deviation of 29 months, when looking at the whole picture. Because of this large standard deviation, much more than 15 months can be saved for some women, but also a lot less in others.

For screening every second year, which is the frequency of the NBCSP, tumors are found a little less than 5 months earlier because of screening, with a standard deviation of about 20 months. Screening every three years only allows for 1.5 months earlier detection, with a standard deviation of 19.5 months.

When screening is even less frequent, tumors are, on average, found earlier clinically than at screening.

It is interesting to see these results of time saved by screening in the light of the results of the growth rate κ. The values of κ in the tumors found at screening were, with two exceptions in the age group 60 - 69, smaller than in tumors found clinically. Especially in younger women, the difference in growth rate is large. This means, that the tumors found at screening grow slower than tumors found clinically, which is natural to assume. Considering that the tumors found at screening are slow growing, and that the time saved, on average, by screening every two years like in the NBCSP is only about 5 months, there is reason to further investigate this matter.

Part V
Bibliography

References

[1] H. Weedon-Fekjær, B. H. Lindqvist, L. J. Vatten, O. O. Aalen and S. Tretli. "Breast cancer tumor growth estimated through mammography screening data", *Breast Cancer Research*, 2008, 10:R41.

[2] Cancer Research UK, "Cancer Worldwide - The Global Picture", http://info.cancerresearchuk.org/cancerstats/geographic/world/
Update: 30. October 2008

[3] Cancer Research UK, "Commonly diagnosed cancers worldwide", http://info.cancerresearchuk.org/cancerstats/geographic/world/commoncancers/
Update: 19. April 2005

[4] J. A. Spratt, D. von Fournier, J. S. Spratt and E. E. Weber. "Decelerating growth and human breast cancer", *Cancer*, 1993, 71:2013-2019.

[5] Norsk Elektronisk Legehåndbok,
http://www.legehandboka.no/asp/document.asp?id=1435
Update: 24 july 2008

[6] Oncolex (encyclopedia of cancer), "Brystkreft",
http://www.oncolex.no/Bryst.aspx/
Update: 10/09/2008.

[7] Cancer Registry of Norway, "Cancer in Norway 2006", http://www.kreftregisteret.no/no/Generelt/Publikasjoner/Cancer-in-Norway/Cancer-in-Norway-2006/, 2007

[8] J. E. Rossouw, G. L. Anderson, R. L. Prentice, A. Z. LaCroix, C. Kooperberg et al. "Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial", *Journal of the American Medical Association*, 2002, 288:321-333.

[9] H. H. Chen, S. W. Duffy and L. Tabár. "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.

[10] H. H. Chen, S. W. Duffy, L. Tabár and N. E. Day. "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.

[11] S. W. Duffy, H. H. Chen, L. Tabar and N. E. Day. "Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase", *Statistics in Medicine*, 1995, 14(14):1531-1543.

[12] S. Y. G. L. Tan , G. J. van Oortmarssen , H.J. de Koning, R. Boer and J. D. F. Habbema. "The MISCAN-Fadia Continuous Tumor Growth Model for Breast Cancer", *Journal of the National Cancer Institute Monographs*, 2006, No. 36.

[13] P. Skekhan, "Cell growth, tissue neogenesis, and neoplastic transformation", in "Growth, cancer, and the cell cycle" edited by P. Shekhan and S. J. Friedman. *Humana Press*, 1984, p.323-45.

[14] Berit Damtjernhaug. "Er mammografiscreening nyttig- og for hvilke aldersgrupper?", *Kirurgen*, 2008, nr.1.

Part VI
Appendix

7 R-code

All codes are stated for the age group 'all', as the code for the other age group is similar.

7.1 Initiation of the simulation

```
### Initiating population ###
n = 25000
time_tot = 50

time = seq(from=0, to=time_tot, by=0.0833333)
months = length(time)

screen_years = 25

### Cancer tumor growth ###
tumorgrowth = matrix(data=NA, nrow=months, ncol=n)
tumorgrowth_det = matrix(data=NA, nrow=months, ncol=n)

tumor = rep(NA, times=months)
lambda = 2.166667e-08

max_time = rnorm(n=n, mean = 240, sd=30)

alfa1 = 1.07
alfa2 = 1.31

k = rlnorm(n=n, meanlog=alfa1, sdlog=alfa2)
k_max = 100

det_diam = rgamma(n=n, shape=2.958, scale = 9.076)
det_value = rep(NA, times=n)

### Other parameters ###
vmax = 4/3 * pi * (128/2)^3
vcell= 1*10^(-6)

beta2 = 6.51
beta1 = 1.47

y <- matrix(rep(1:months,n),months)/12
```

```
### Simulation (age group all) ###

### Making sure that kappa is not unrealistically large ###
for(i in 1:n) {
while (k[i] > k_max) (k[i] = rlnorm(n=1, meanlog=alfa1, sdlog=alfa2))
}

### A simulation of all potential tumors ###
for(i in 1:n) {

  for(t in 1:months){
  tumor[t]= 2*((3*vmax)/(4*pi*((1 + ((vmax/vcell)^0.25 - 1)*
  exp(-0.25*k[i]*time[t]))^4)))^(1/3)

    }
tumorgrowth[,i] = tumor
}

tumorgrowth_det = tumorgrowth

### Checking where all the tumors will be detected clinically ###
for (i in 1:n) {
    for(t in 1:months){
         if (tumorgrowth_det[t,i] > det_diam[i])
         (det_value[i] = tumorgrowth_det[t,i])
         & (tumorgrowth_det[(t:months),i] = 0)
    }
}

### Taking out "old" tumors ###
for (i in 1:n) {
    for (t in max_time[i]:months+1) {
    tumorgrowth_det[t,i] = 0
    }
}
```

7.2 Screening

```
### Screening frequency ###
a = 1
j = months - (screen_years*12 +1)
screen = seq(from= j+(a*12), to= months-(5*12+1) , by=a*12)

### Selection of cancer and cancer growth ###
pop = rep(0, times=n)
cancer = rep(0, times=n)

growth_no_screen = matrix(data=0, nrow=months, ncol=n)
growth_with_screen = matrix(data=NA, nrow=months, ncol=n)
tumor_dist = matrix(data=NA, nrow=months, ncol=n)
```

```
screen_det = rep(0, times=n)
sts = rep(NA, times=n)

### Some people get cancer each month ###
for(t in 1:months){
cancer = rpois(n=n, lambda=(n*lambda))
    for (i in 1:n) {
    if (cancer[i] == 1) (pop[i] = 1) & (growth_no_screen[t:months,i]
    = tumorgrowth_det[1:(length(t:months)),i])
    }
}

pop_clin = pop

### Screening ###

### Assuming the first is screening at time j ###

for (i in 1:n) {
sts[i] = exp((growth_no_screen[j,i] - beta2)/beta1) /
( 1 + exp((growth_no_screen[j,i] - beta2)/beta1))
}

for (i in 1:n) {
if (sts[i] < 0.016489) (sts[i] = 0)
}

### Detection according to a binomial distribution (1, sts[i]) ###

tumor_detected = rep(NA, times=n)

for(i in 1:n) {
tumor_detected[i] = rbinom(n=1, size=1, prob=sts[i])
}

### Removing the tumors that were detected at screening at time j ###

for(i in 1:n) {
if (tumor_detected[i] == 1) (growth_with_screen[(j+1),i] = 0) & (pop[i] = 0)
& (tumor_dist[j,i] = growth_no_screen[j,i]) & (screen_det[i] = 1)
    else (growth_with_screen[(j+1),i] = growth_no_screen[(j+1),i])
}

growth_with_screen[0:j,] = growth_no_screen[0:j,]

for (i in 1:n) {
if (pop[i] == 1) (growth_with_screen[(j+2):screen[1],i] =
    growth_no_screen[(j+2):screen[1],i])

    else (growth_with_screen[(j+1):months,i] = 0)
```

```
}


for (i in 1:n) {
sts[i] = exp((growth_no_screen[j,i] - beta2)/beta1) /
( 1 + exp((growth_no_screen[j,i] - beta2)/beta1))
}

for (i in 1:n) {
if (sts[i] < 0.016489) (sts[i] = 0)
}

tumor_detected = rep(NA, times=n)

for(i in 1:n) {
tumor_detected[i] = rbinom(n=1, size=1, prob=sts[i])
}

### How many people have cancer before screening? ###

cases_before[1] = sum(pop)
total_volume_before[1] = sum(growth_no_screen[j,])

for(i in 1:n) {
if (tumor_detected[i] == 1) (growth_with_screen[(j+1),i] = 0) & (pop[i] = 0)
& (tumor_dist[j,i] = growth_no_screen[j,i]) & (screen_det[i] = 1)

    else (growth_with_screen[(j+1),i] = growth_no_screen[(j+1),i])
}

### How many people have cancer after screening? ###

growth_with_screen[0:j,] = growth_no_screen[0:j,]

for (i in 1:n) {
if (pop[i] == 1) (growth_with_screen[(j+2):screen[1],i] =
growth_no_screen[(j+2):screen[1],i])

    else (growth_with_screen[(j+1):months,i] = 0)
}

cases_after[1] = sum(pop)
total_volume_after[1] = sum(growth_with_screen[j+1,])

gc()

### Screening (for all other then the first and last) ###

for (v in 1:(length(screen)-1)) {
sts = rep(NA, times=n)
```

```
  for (i in 1:n) {
  sts[i] = exp((growth_with_screen[screen[v],i] - beta2)/beta1) /
     ( 1 + exp((growth_with_screen[screen[v],i] - beta2)/beta1))
     }

        for (i in 1:n) {
        if (sts[i] < 0.016489) (sts[i] = 0)
        }

                tumor_detected = rep(NA, times=n)

                for(i in 1:n) {
                tumor_detected[i] = rbinom(n=1, size=1, prob=sts[i])
                }

### Removing the tumors that were detected at screening at time screen[v] ###
  for(i in 1:n) {

   if (tumor_detected[i] == 1) (growth_with_screen[(screen[v]+1),i] = 0) & (pop[i] = 0)
   & (tumor_dist[screen[v],i] = growth_with_screen[screen[v],i]) & (screen_det[i] = 1)

         else (growth_with_screen[(screen[v]+1),i] = growth_no_screen[(screen[v]+1),i])
   }


### How many people have cancer after screening? ##
  for (i in 1:n) {

   if (pop[i] == 1) (growth_with_screen[(screen[v]+2):screen[v+1],i] =
   growth_no_screen[(screen[v]+2):screen[v+1],i])

          else (growth_with_screen[(screen[v]+1):months,i] = 0)
   }
}

### What happens after the last screening? ###
sts = rep(NA, times=n)

for (i in 1:n) {
sts[i] = exp((growth_with_screen[screen[length(screen)],i] - beta2)/beta1) /
( 1 + exp((growth_with_screen[screen[length(screen)],i] - beta2)/beta1))
}

        for (i in 1:n) {
        if (sts[i] < 0.016489) (sts[i] = 0)
        }


 tumor_detected = rep(NA, times=n)

  for(i in 1:n) {
```

```
  tumor_detected[i] = rbinom(n=1, size=1, prob=sts[i])
  }


### Removing the tumors that were detected at screening at time screen[v] ###

   for(i in 1:n) {

   if (tumor_detected[i] == 1) (growth_with_screen[(screen[length(screen)]+1),i] = 0)
   & (pop[i] = 0) & (tumor_dist[screen[length(screen)],i] =
   growth_with_screen[screen[length(screen)],i]) & (screen_det[i] = 1)

        else (growth_with_screen[(screen[length(screen)]+1),i] =
        rowth_no_screen[(screen[length(screen)]+1),i])
    }

### How many people have cancer after screening? ###

  for (i in 1:n) {

    if (pop[i] == 1) (growth_with_screen[(screen[length(screen)]+2):months,i] =

        else (growth_with_screen[((screen[length(screen)]+1):months),i] = 0)
   }
```

7.3 Calculations of time saved

```
### How much time is saved because of screening in the original simulation? ###
growth_time = rep(0, times=n)      ng

        for (i in 1:n) {
                if (screen_det[i] == 1)

        (for (t in 1:months) { if (tumorgrowth[t,i] > 0)
           ( if (tumorgrowth[t,i] < det_diam[i])
           (growth_time[i] = growth_time[i]+1)) } )
        }

tumor_dist2 = rep(NA, times=n)

 for (i in 1:n) {
    xx = rep(NA, times=months)
       if (screen_det[i] == 1)  (xx = tumor_dist[,i]) & (xx = xx[!is.na(xx)])
        & (tumor_dist2[i] = xx)
    }

growth_time2 = rep(0, times=n)
```

```
 for (i in 1:n) {
    if (screen_det[i] == 1)

      (for (t in 1:months) { if (tumorgrowth[t,i] > 0) ( if (tumorgrowth[t,i]
       <= tumor_dist2[i]) (growth_time2[i] = growth_time2[i]+1)) } )
 }

growth_time[is.na(growth_time)] = 0

saved_time1 = growth_time - growth_time2

### How much time is saved for different screening intervals,
### compared to clinical finding ###

tumorgrowth_new = matrix(data=0, nrow=months, ncol=n)

for (i in 1:n) {

     t = runif(1, 1, 600)
     tumorgrowth_new[t:months,i] = tumorgrowth[1:length(t:months),i]
}

### Screening every year ###
new_time_total = 40
new_time = seq(from=0, to=new_time_total, by=0.0833333333)
new_months = length(new_time)

a = 1
new_screen = seq(from= a*12, to= new_months , by= a*12 )
new_sts = matrix(data=NA, nrow=new_months, ncol=n)

new_tumor_detected = matrix(data=0, nrow=new_months+1, ncol=n)
new_det = rep(0, times=n)

for (i in 1:n) {

  for (t in new_screen) {
         new_sts[t,i] = exp((tumorgrowth_new[t,i] - beta2)/beta1) /
         ( 1 + exp((tumorgrowth_new[t,i] - beta2)/beta1))

         new_tumor_detected[t,i] = rbinom(n=1, size=1, prob=new_sts[t,i])
  }
}

for (i in 1:n) {

  for (t in new_screen) {
      if (new_sts[t,i] < 0.016489) (new_sts[t,i] = 0)
  }
}
```

```
for (i in 1:n) {

  for (t in new_screen) {
       if (new_tumor_detected[t,i] == 1)
       (new_tumor_detected[t:new_months,i] = 1) & (new_det[i] = 1)
  }
}

growth_time3 = rep(0, times=n)

for (i in 1:n) {

   for (t in 1:new_months) {  if (tumorgrowth_new[t,i] < det_diam[i])
   (growth_time3[i] = growth_time3[i]+1)
   }
}

growth_time4 = rep(0, times=n)   # Time until tumors found at screening

for (i in 1:n) {

   for (t in 1:new_months) { if (new_tumor_detected[t,i] == 0)
            (growth_time4[i] = growth_time4[i]+1)
   }
}

saved_time_int1 = growth_time3 - growth_time4
```

7.4 Calculations of κ

```
### Finding the distribution of kappa in tumors that are found clinically ###

total_pop = (pop_clin - screen_det)
kappa_clin = rep(NA, times=n)

for (i in 1:n) {
    if (total_pop[i] == 1) (kappa_clin[i] = k[i])
}

kappa_clin1 = kappa_clin[!is.na(kappa_clin)]

### Finding the distribution of kappa in tumors that are found at screening ###

kappa_screen = rep(NA, times=n)

for (i in 1:n) {
    if (screen_det[i] == 1) (kappa_screen[i] = k[i])
}

kappa_screen1 = kappa_screen[!is.na(kappa_screen)]
```