Estimation of Overdiagnosis in Breast Cancer Screening

I. Summary

The benefit of breast cancer (BC) screening with mammography has been demonstrated by randomized controlled trials (RCTs). However, there is a growing concern about the harms of screening, in particular the problem of overdiagnosis. It occurs when screening detects a breast cancer that would not have presented clinically in the woman's lifetime in the absence of screening. Overdiagnosis results in unnecessary biopsies, surgery and overtreatment with psycho-social consequences. Therefore accurately quantifying the frequency of overdiagnosis is important for informed decision making and policy development. Unfortunately, at diagnosis of an asymptomatic breast cancer it is impossible to determine if the breast cancer will become clinically apparent during the women's' lifetime. Thus measurement of the frequency of overdiagnosis is not straightforward.

The classical way for estimation of overdiagnosis is to compare the cumulative incidence of breast cancer between screened and unscreened women which may work in RCTs with a certain design. In nationwide population-based service screening programs there is no obvious contemporaneous control group for evaluation. In addition, published estimates of overdiagnosis vary considerably, from less than 1% and up to 54%. The discrepancy can probably be explained by different study designs, choice of control groups, follow-up time, and evaluation methods. We have developed a non-homogenous multi-state model for estimation of the frequency of overdiagnosis. The great advantage of this method is that no control group is needed and overdiagnosis can be estimated also for a short and recent time period. In a simulation study we have showed that this method results in accurate estimates.

The RCTs on mammography screening are old, the mammography technique has developed, e.g from analogue to digital, and become more sensitive. Furthermore a new screening tool, i.e. tomosynthesis (3-dimensional mammography), with even higher sensitivity may soon be introduced in routine screening. Since mammography screening is known to be effective, new RCTs with an unscreened control group would be unethical. Thus methods like multi-state models, which do not require an unscreened control group, are not just suitable but also necessary for evaluation of the frequency of overdiagnosis in the future. It is very important that the models are robust and give correct estimates. We want to further develop and validate our model and apply it on national mammography screening data from Norway, Demark, Finland and Sweden.

II. Purpose and aim

The focus of the project covers three parts.

1. **Further development of the multi-state model with respect to**
    a. Increased flexibility to model age, period or cohort effect.
    b. Efficient algorithms for estimation

2. **Evaluation of overdiagnosis in current Nordic mammography screening data**
    a. In Norway (start 1996)
    b. In Finland (start 1987)
    c. In Denmark (20% start from 1991 and remaining 80% from 2008)

d. In Sweden (start 1986)

3. Estimation of overdiagnosis in screening with tomosynthesis
In three recent trials in Norway

III. Background

The benefit of BC screening with mammography in terms of BC mortality reduction has been demonstrated by the RCTs. In a meta-analysis of 11 RCTs, 20% mortality reduction was found.(1) However, there is a growing concern about the harms of screening, in particular the problem of overdiagnosis. Overdiagnosis is defined as BC detected at screening that would not have surfaced clinically in the women's lifetime in the absence of screening. However, at diagnosis of an asymptomatic BC it is currently impossible to determine if the BC will become clinically apparent in the woman's lifetime. Therefore, monitoring the frequency of overdiagnosis is not straightforward and can only be estimated on population level.(2)

The preferred situation for estimation of overdiagnosis is to compare the cumulative incidence of BC between women invited to screening and a control group in a RCT with a stop-screen design, i.e. no screening is offered after the RCT is closed. Follow-up of the study population should continue after the RCT was closed in order to allow catch-up of women in the unscreened control group to compensate for cases diagnosed earlier in the screening group. The excess number of BCs between the two groups provides a measure of overdiagnosis. This method is called the cumulative incidence method (CIM) or the excess incidence approach.(3, 4) The majority of the performed trials were not designed to allow for this follow-up and in the meta-analysis only three of the RCTs were included to estimate overdiagnosis.(1) A recent review of the observational studies showed a huge variation in the estimates (1% - 54%).(5) This discrepancy might result from different study designs, estimation methods and methods used to adjust for lead time, which is the length of time by which screening advances the diagnosis compared with absence of screening.(4, 5)

The control group in observational studies is often obtained by extrapolation of the incidence trends from a prescreening period (historical control) or by selection of subjects who lived in the geographic areas without screening program over the same period (geographic control). However, period effect (e.g. increasing incidence) or cohort effect might cause that the control group is not representative for the screened group in absence of screening.(6)

An alternative approach in which an external control group is not needed is to estimate overdiagnosis through a multi-state model (MSM) where the natural progression of BC is estimated by using individual screening histories. Several MSMs have been developed to quantify overdiagnosis but the common assumption was that the incidence rate of BC was constant over age.(7-10) Using a simulated screening scenario we observed that using constant rates in MSMs resulted in poor goodness of fit to the data and an underestimation of the overdiagnosis. We have, in our research group, developed a non-homogenous multi-state model which can handle the age-specific transitions rates e.g. incidence. In a recent simulation study our non-homogeneous MSM gave similar estimates on overdiagnosis as the CIM.(11) We believe that a similar approach can be applied for evaluation of overdiagnosis in service screening programmes.

IV. Project description

We have developed a MSM to describe the natural history of BC. The model included four states denoted by $X(t)$: free of BC (state 1), progressive preclinical screen-detectable phase (PCDP) (state 2), clinical phase (CP) (state 3) and non-progressive PCDP, (state 4). In order to distinguish the collected information from the invisible latent states we used three observed states denoted by $Y(t)$, including normal (negative) mammographic finding (observed state 1), screen-detected case (observed state 2), and clinically detected case (observed state 3). The test sensitivity (S) is defined as the probability of a positive screening finding among those who are in the PCDP states. Figure 1 illustrates the possible latent states, transitions denoted by arrows and the probabilities (1, S, and 1-S) to be observed in the different states given the latent state.

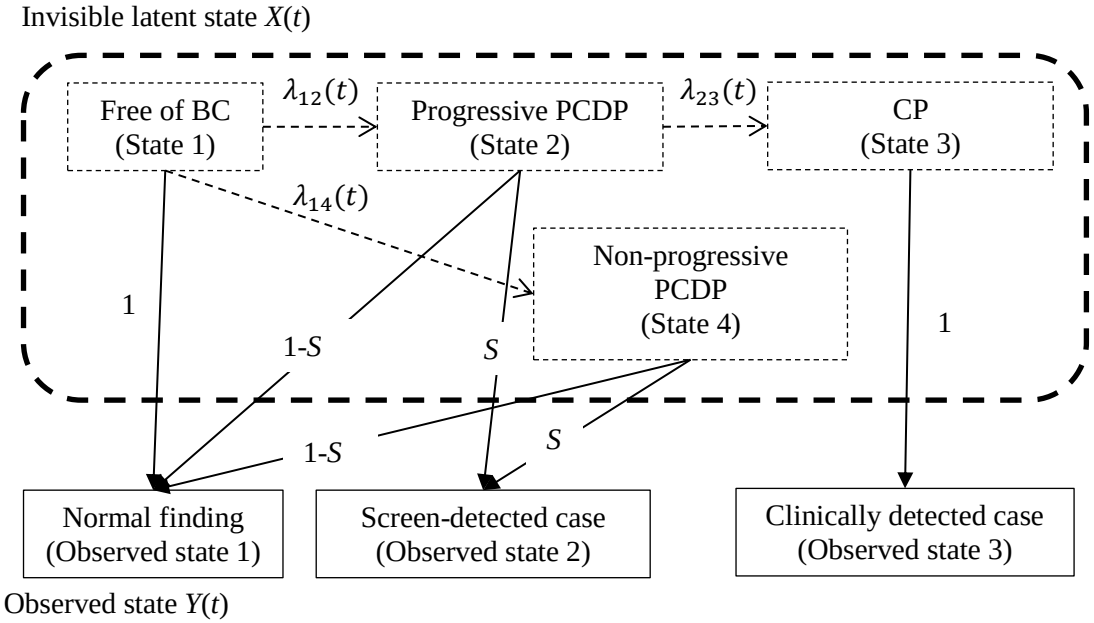

Figure 1. The natural history and observed status of breast cancer

The latent process $\{X(t)\}$ is assumed to follow the Markov assumption, which implies that the future process given the current state is independent of the earlier history of the process. The transition probabilities are represented by $P_{ij}(s,t) = \Pr\left(X(t) = j \mid X(s) = i\right)$ for $i, j = 1,...,4$ and $0 \le s \le t$. Following Cox and Miller's continuous-time Markov process, the transition rates are represented by the derivatives $\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t, t+\Delta t)}{\Delta t}$.(12)

Individual probabilities for the observed findings of $Y(t)$ are expressed as formulas of $P_{ij}(s,t)$ for $i, j = 1,...,4$, $i \le j$ and s, t are woman's age at screening except for clinically detected cases where t is age at diagnosis. The overall likelihood function is the product over all individuals. The quasi-Newton algorithm is used to get the maximum likelihood estimates (MLEs) from the log-likelihood function. Approximated standard errors of the parameter estimates are obtained from the inverse of the matrix of second derivatives (Hessian matrix) of the maximized log-likelihood function.

Further development of the multi-state model (aim 1a)

We plan to develop a more flexible MSM which can accommodate different situations including age and period effects. Specifically,

1. Age effect
   (1) Age-dependent transition rates
   The BC incidence increases with age. Two modelling approaches as below will be used.
   a. Piecewise-constant transition rates
   We consider a piecewise-constant transition rate matrix in the disjointed intervals determined by a set of cut-points $0 = a_0 < a_1 < ... < a_R = \infty$, then the transition rates can be expressed by $\lambda_{ij}(t) = \lambda_{ij}^{(l)}$ for $a_{l-1} \le t < a_l$. (13)
   b. Weibull type of transition rates
   We plan to introduce a parametric Weibull distribution to describe the age-dependent transition rates. It can be expressed by $\lambda_{ij}(t) = \frac{\alpha_{ij}}{\beta_{ij}} t^{\alpha_{ij}-1}$, where α_{ij} is the shape parameter and β_{ij} is the scale parameter. The transition rate increases with age when $\alpha_{ij} > 1$ and decreases over time when $\alpha_{ij} < 1$. A value of $\alpha_{ij} = 1$ indicates that the transition rate is constant over time.
   (2) Sensitivity
   Dense tissue has generally been associated with younger age and high breast density will decrease the sensitivity of mammography. We plan to quantify the sensitivity for different age groups separately.
2. Period effect
   (1) Changing incidence rates
   The incidence of BC has increased over time in many countries.(14) We plan to use proportional hazards to consider the period effect on transition rates.
   (2) Sensitivity
   The sensitivity of mammography screening may have improved by the developments in the practice of screening and introducing new techniques. We plan to quantify the sensitivity for different time periods separately.

Efficient algorithm for estimation (aim 1b)

The large amount of data from the individual screening histories and the complicated likelihood function results in a time-consuming estimation process. We are going to develop an efficient algorithm to handle the likelihood function and perform the estimation procedure through the resource in High Performance Computing Center North (HPC2N) (15) to reduce the computation time.

Evaluation of overdiagnosis in current Nordic mammography screening data (aim 2)

We are going to collaborate with researchers from Norway, Finland and Denmark where national registers on the mammography screening program are available. Data from the three countries will be checked and cleaned and linked to the national cancer registers. Estimation of overdiagnosis will then be made using the optimal MSM. A nationwide quality register on mammography screening is under development in Sweden and data will be used for estimation as soon they are available. The screening programs are described in Table 1.

Estimation of overdiagnosis in tomosynthesis (aim 3)

Digital breast tomosynthesis is a newly developed three-dimensional imaging technique that has the potential to improve the accuracy of mammography. We plan to collaborate with the researchers in Norway by using data from the OTST (Oslo Tomosynthesis Screening Trial), the OVVV-study (Oslo, Vestre Viken og Vestfold), and the TOBE-trial (Tomosynthesis in Bergen) to evaluate the overdiagnosis from tomosynthesis.(16-18) In total over 55,000 women are screened in the above trials.

Table.1 Description of the screening programs and population in the four countries

Mammography screening program	**National/ regional**	**Start first areas**	**Start last areas**	**Age limits**	**Total population 2013 (million)**
Sweden	National	1986	1997	40-74* current 50-69* minimum	9.6
Finland	National	1987	1992	50-69 current 50-59 minimum	5.4
Denmark 20% of the country	Regional	1991	1995	50-69	1.1
Norway	National	1996	2004	50-69	5.2
Denmark remaining 80% of the country	National	2008	2010	50-69	4.5

* corresponding to 43% and 24% of the female population in 2014, respectively

V. Preliminary results

Model performance on piecewise-constant transition rate

We conducted a simulation study to verify the validity of using MSM on the estimation of overdiagnosis. The invisible latent disease progression of BC and the observed disease states for one-million women with biennial screening regime in women aged 50 to 69 were simulated. The BC incidence rate from the cancer registry and the published information on sojourn times (maximum individual lead time), transition rates to non-progressive PCDP and test sensitivity from trials were used to determine the parameters.(19, 20) We developed a non-homogenous MSM model with piecewise-constant transition rates and constructed the likelihood function for estimation of the parameters. 100 replications of simulation with one million women each were done to investigate the performance of the MSM and CIM. The results are summarised in Figure 2. The mean value of the true frequency of overdiagnosis was 12.5% and the average estimates by the CIM and the MSM were 12.9% (interquartile range 2.5%) and 13.4% (interquartile range 2.2%), respectively. The MSM had a somewhat larger bias than CIM but the variation of the estimates was smaller.(11)

Further development of the multi-state models

An alternative to capture the age-dependent BC incidence rate, instead of piecewise-constant exponential distribution, is a parametric distribution. We used the Weibull distribution to model the age-specific BC in Sweden in years 1981 - 1985. The observed incidence (dots) and fitted Weibull hazard rate (curve) are shown in Figure 3.

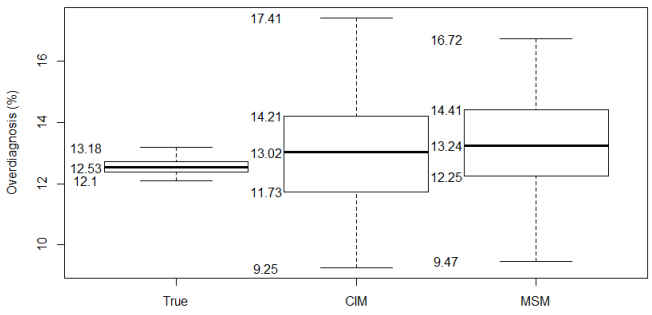

Figure 2. The boxplot of overdiagnosis estimated by CIM and MSM compared with the true values in the simulation study.

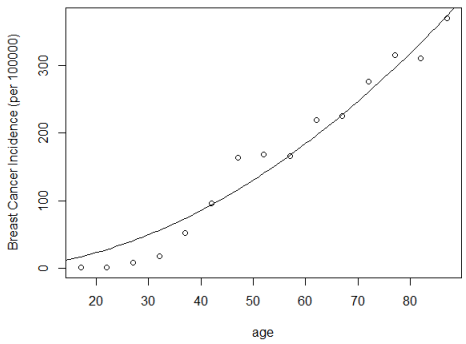

Figure 3. The observed incidence rate of breast cancer in Sweden in 1981-1985 (dots) and fitted results (curve)

Evaluation of overdiagnosis in current Nordic mammography screening data

We have retrieved individual screening data from Stockholm and Norrbotten county in Sweden. Screening characteristics and outcomes are summarised in Table 2. In an ongoing project we are using the MSM model developed above to estimate overdiagnosis in these two counties.

Table 2. Description of the retrieved screening information from the service screening in Stockholm and Norrbotten, Sweden

Screening information	**Stockholm**	**Norrbotten**
Target population (age)	50-69*	40-74
Year of start screening	1989	1989
Enrolled birth cohort	1920-1959	1915-1972
End of follow-up	Until age 76 or year 2014	Until age 76 or year 2012
Study population	418,294	94,537
Invitation and participation		
Total number of invitation	2,333,863	638,810
Total number of mammography	1,696,273	538,858
Participation rate	72.68%	84.35%
Total number of recall	42,030	10,486
Recalled rate	2.48%	1.95%
Mode of detection		
Prevalent screen-detected cancer	1,780 (10.11%)	249 (8.64%)
Subsequent screen-detected cancer	6,437 (36.54%)	1,375 (47.73%)
Interval cancers	3,633 (20.62%)	598 (20.76%)
Cancer in non-participants	2,809 (15.95%)	198 (6.87%)
Cancer in uninvited women	2,956 (16.78%)	461 (16.00%)
Total breast cancer	17,615 (100%)	2,881 (100%)

* In 2005 and following years it was expanded to 40-69 years and in 2012 to 40-74 years

VI. Team members and collaborators

The researchers in this project will include

1. **Håkan Jonsson,** principal investigator, associate professor, Umeå University, Sweden
2. **Wendy Yi-Ying Wu,** PhD, statistician, Umeå University, Sweden
3. **Lennarth Nyström**, associate professor, Umeå University, Sweden
4. **Solveig Hofvind,** professor, Cancer Registry of Norway, Oslo and Akershus University College, Norway
5. **Søren Nymand Lophaven**, senior statistician, University of Copenhagen, Denmark
6. **Tytti Sarkeala**, director of screening, PhD, Mass Screening Registry/Finnish Cancer Registry, Finland
7. **Sven Törnberg,** associate professor, Stockholm-Gotland Regional Cancer Centre, Sweden
8. **Sirpa Heinävaara**, statistician, associate professor, Mass Screening Registry/Finnish Cancer Registry

VII. Ethical consideration

In collaboration with group members in respective country we will apply for ethical approval and other necessary permissions to use nationwide individual but non-identifiable data on screening and follow-up of breast cancer diagnosis and death in the national registers (Table 1). We already have an approval for Stockholm and north Sweden.

VIII. Significance

Breast cancer is the most common female cancer in the Western world. The target group for BC screening in Sweden, Norway, Denmark and Finland is roughly 4 million women which with biannual screening means 2 million women invited to screening every year. Overdiagnosis and its consequences is the most important harm in BC screening. There is no consensus how to estimate its frequency in service screening because of the methodological difficulties. Studies with better methods are thus needed to quantify overdiagnosis and to motivate its meaningful estimation. The great advantages of multi-state model methods are that an unscreened control group is not needed and that overdiagnosis can be estimated also for a recent period. Hopefully estimating overdiagnosis can become a task of the monitoring of the screening programs in the future. We aim to provide new development on the estimation methods, add valuable evidence about methodological pros and cons, and to actually estimate overdiagnosis at mammography screening in four of the Nordic countries. The developed method will also be used for evaluation of overdiagnosis in screening with the new test, tomosynthesis.

IX. References

1. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380(9855):1778-86.
2. Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J Med Screen. 2012;19 Suppl 1:5-13.
3. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol. 2007;8(12):1129-38.
4. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med. 2013;158(11):831-8.

5. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56.
6. Ripping TM, Verbeek AL, Ten Haaf K, van Ravesteyn NT, Broeders MJ. Extrapolation of pre-screening trends: Impact of assumptions on overdiagnosis estimates by mammographic screening. Cancer Epidemiol. 2016;42:147-53.
7. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res. 2005;7(6):258-65.
8. Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J. 2006;12(4):338-42.
9. Gunsoy NB, Garcia-Closas M, Moss SM. Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40 to 49 in the United Kingdom. Breast Cancer Res. 2012;14(6):R152.
10. Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer. 2003;39(12):1746-54.
11. Wu W, Nyström L, Jonsson H. Estimation of overdiagnosis in breast cancer screening using a non-homogenous multi-state model: a simulation study. Submitted.
12. Cox DR, Miller HD. The Theory of Stochasic Processes. London: Chapman & Hall; 1965.
13. Ocana-Riola R. Non-homogeneous Markov processes for biomedical data analysis. Biom J. 2005;47(3):369-76.
14. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev. 2015;24(10):1495-506.
15. High performance computing center north (HPC2N) https://www.hpc2n.umu.se/: Umeå University. Available from: https://www.hpc2n.umu.se/.
16. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, Izadi M, Jebsen IN, Jahr G, Krager M, Hofvind S. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol. 2013;23(8):2061-71.
17. Roth Hoff S, Iversen B, Lømo J, Pedersen K, Skaane P. Tomosynthesis in mammography screening (In Norwegian) Oslo: Kreftregisteret; 2015. Available from: https://www.kreftregisteret.no/globalassets/publikasjoner-og-rapporter/mammografiprogrammet/rapport-tomosyntese.pdf.
18. The Tomosynthesis Trial in Bergen (TOBE) ClinicalTrials.gov2016. Available from: https://clinicaltrials.gov/ct2/show/study/NCT02835625?term=tomosynthesis&rank=1.
19. Duffy SW, Day NE, Tabar L, Chen HH, Smith TC. Markov models of breast tumor progression: some age-specific results. J Natl Cancer Inst Monogr. 1997(22):93-7.
20. Moss S, Waller M, Anderson TJ, Cuckle H. Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. Br J Cancer. 2005;92(5):955-60.