

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/207065>

Please be advised that this information was generated on 2021-06-19 and may be subject to change.

Test sensitivity of mammography and mean sojourn time over 40 years of breast cancer screening in Nijmegen (The Netherlands)

J Med Screen
2019, Vol. 26(3) 147–153
© The Author(s) 2018



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0969141318814869
journals.sagepub.com/home/msc



AMWM Aarts¹, SW Duffy², SME Geurts³, DP Vulkan², JDM Otten¹,
C-Y Hsu⁴, THH Chen⁴, ALM Verbeek¹ and MJM Broeders^{1,5}

Abstract

Objectives: We investigated whether changes in mammographic technique and screening policy have improved mammographic sensitivity, and elongated the mean sojourn time, since the introduction of biennial breast cancer screening in Nijmegen, the Netherlands, in 1975.

Methods: Maximum likelihood estimation, non-linear regression, and Markov Chain Monte Carlo simulation were used to estimate test sensitivity, mean sojourn time, and underlying breast cancer incidence in four time periods, covering 40 years of breast cancer screening in Nijmegen (1975–2012).

Results: Maximum likelihood estimation generated an estimated test sensitivity of approximately 90% and a mean sojourn time around three years, while the estimates based on non-linear regression and Markov Chain Monte Carlo simulation were 80% and four years, respectively. All three methods estimated a rise in the underlying breast cancer incidence over time, with approximately one case more per 1000 women per year in the final period compared with the first period.

Conclusions: The three methods showed a slightly higher mammographic sensitivity and a longer mean sojourn time in the last period, after the introduction of digital mammography. Estimates were more realistic for the more sophisticated methods, non-linear regression and Markov Chain Monte Carlo simulation, while the simple closed form approximation of maximum likelihood estimation led to rather high estimates for sensitivity in the early periods.

Keywords

Breast cancer screening, mammography, sojourn time, test sensitivity

Date received: 4 July 2018; accepted: 1 November 2018

Introduction

Analyses of the national breast cancer screening program in the Netherlands have shown that breast cancer can be detected early and that breast cancer mortality has decreased since the introduction of screening, while estimates of overdiagnosis are considered acceptable.¹ The start of the Dutch breast cancer screening program in 1989 was preceded by pilot studies in Utrecht and Nijmegen.^{2,3} Since the introduction of biennial breast cancer screening in Nijmegen in 1975, the mammographic technique has improved, and has recently changed from analogue to digital. The hypothesis is that these technological advances, as well as other changes within the program (e.g. to increase the referral rate) led to an improved test sensitivity of mammography, and an elongation of the mean sojourn time. Sojourn time can be defined as the duration of the preclinical screen-detectable phase, i.e. the period during which a woman is asymptomatic but

the breast cancer is detectable by mammography.⁴ Test sensitivity is the probability that a woman with asymptomatic breast cancer undergoing mammographic screening during the preclinical detectable phase will have her

¹Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

²Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

³Department of Medical Oncology, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

⁴Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei

⁵Dutch Expert Centre for Screening, Nijmegen, The Netherlands

Corresponding author:

AMWM Aarts, Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands.

Email: Anne.Aarts@Radboudumc.nl

breast cancer detected by the test.⁴ As the effectiveness of breast cancer screening crucially depends on test sensitivity and mean sojourn time, these measures were estimated when the breast cancer screening program was introduced, but the estimates have not been updated following the technological advances in mammography and policy changes in the program.⁵

The challenge of quantifying mean sojourn time is that it is not directly observable. It can, however, be estimated. The easiest method is to use tumour volume doubling times as a proxy, as mean sojourn time can be seen as a measure of tumour growth.^{6,7} Several methods, some simple and some complex, have been developed for the estimation of mean sojourn time based on screening frequency data and breast cancer prevalence and incidence data, which are easily obtainable. Methods described in the current literature include simple methods based on the ratio of prevalence of the disease at first screening examination related to the expected annual incidence rate,⁸ parametric models assuming a specific distribution for sojourn time,^{9,10} nonparametric methods with time split into discrete intervals,¹¹ simulation models with patient-level data,¹² and Markov models simulating the natural history of chronic diseases.¹³ As there is no gold standard to estimate mean sojourn time, the reliability of all these estimation methods remains unclear. We therefore applied three different estimation methods, to investigate whether they gave similar results which can still be incorrect. As our interest was in measuring the effect of technological advances in mammography and changes in policy over several decades, we wanted to obtain multiple estimates for test sensitivity and mean sojourn time. The most relevant change is the shift from analogue to digital mammography in 2007, and its accompanying higher breast cancer detection rate.¹⁴

In this paper, we investigate whether technological advances in mammography and changes in screening policy have improved the test sensitivity of mammography and elongated the correlated mean sojourn time. We illustrate this by applying three different estimation methods, using almost 40 years of data from the Nijmegen breast cancer screening program.

Methods

A pilot study on biennial breast cancer screening with analogue mammography started in 1975 in the city of Nijmegen, the Netherlands.¹⁵ After promising results from this pilot, and another in Utrecht, the Dutch government decided to implement nationwide biennial breast cancer screening,^{2,3,16} and from 1989 the Nijmegen program became part of the national program. The Dutch government was responsible for the execution and quality control of the program and the training of the radiologists and radiographers. After the implementation period of the national program, the referral rate was 10 per 1000 women screened, with a somewhat disappointing breast cancer detection rate. The result triggered an investigation into

the optimal referral rate in the early 2000s.¹⁷ Based on this ‘optimization study,’ the Dutch Expert Centre for Screening recommended an increase in the referral rate to 20 per 1000 women screened. In 2007, the breast cancer screening program in Nijmegen switched from analogue to digital mammography.¹⁸ This further increased the referral rate to around 25 per 1000 women screened.¹⁹

During the 40 years of breast cancer screening in Nijmegen, women in varying age ranges were invited, but women aged 50–69 were consistently invited. For these women, the following data were collected: invitation for screening, participation in screening, referral for further diagnostic work-up, and diagnosis of a screen-detected cancer or interval cancer (a cancer detected between two consecutive screening rounds). These data were used to estimate test sensitivity of mammography and mean sojourn time. To investigate the effect of changes in mammographic technique and screening policy on these two parameters, the data of the Nijmegen breast cancer screening program (1975–2012) were grouped in four periods: (1) pilot study in Nijmegen (1975–1988); (2) introduction of nationwide breast screening program (1989–2000); (3) publication of study on increasing the Dutch referral rate (2001–2006); and (4) introduction of digital mammography (2007–2012). For each of these periods we obtained estimates of test sensitivity, mean sojourn time, and the underlying breast cancer incidence. Because breast cancer screening in the Netherlands was implemented more than 25 years ago (1989), there was no suitable control group or other reliable estimate of the underlying breast cancer incidence available.

We here describe three methods for estimating test sensitivity of mammography and mean sojourn time based on modeling routine screening outcome data.

Method 1: based on maximum likelihood estimation

Test sensitivity, mean sojourn time, and underlying incidence were estimated based on empirical screening data. To determine the test sensitivity of mammography we calculated the ratio of the number of screen-detected cancers to the number of interval cancers diagnosed in the first year after screening, plus all screen-detected cancers. The assumption was made that interval cancers detected in the first year after screening were missed cancers from the previous screen.²⁰ Let λ_1 be the underlying incidence of pre-clinical disease, λ_2 be the rate of disease progression from pre-clinical to clinical phase, and S the screening test sensitivity (of mammography). Let t be the interval between screens in years. At first screen, the expected proportion of persons found to have cancer would be

$$P = \frac{\lambda_1 S}{\lambda_2}$$

At incident screens, the expected proportion is²¹

$$I = S \left\{ \int_0^t \lambda_1 e^{-\lambda_1 s} e^{-\lambda_2(t-s)} ds + (1-s)e^{-\lambda_2 t} \int_0^t \lambda_1 e^{-\lambda_1 s} e^{-\lambda_2(t-s)} ds \right\}$$

The first term within the brackets is the probability of cancers newly arising in the preclinical phase since the last screen and not progressing to a clinical disease before the next screen. The second term is the probability of preclinical cancers missed at the previous screen which have not progressed to the clinical phase within the screening interval. There are two simplifying approximations here. The first is that the first term of the formula applies if the previous screen was also a subsequent (not a first) screen. Literature has shown that interval cancer rates after a subsequent screen are similar to interval cancer rates after a first screen.²² Applying this universally, regardless of the status of the previous screen, is therefore arguably a reasonable approximation. The second term is the absence of terms for cancers missed at screens before the last screen. Here, we assume that if a cancer is missed at a screen, it will progress to a clinical disease in the subsequent interval, be detected at the subsequent screen, or progress to a clinical disease in the interval following the subsequent screen. This is an approximation; however, we assume that the probability of missed cancer at two successive screens is relatively small.²³

The above formula with calculus then solves to

$$I = S \left\{ \frac{\lambda_1(e^{-\lambda_1 t} - e^{-\lambda_2 t})(1 + (1 - S)e^{-\lambda_2 t})}{(\lambda_2 - \lambda_1)} \right\}$$

The expected proportion of screen-negative subjects having a clinical interval cancer before the next screen (making the same approximations as for subsequent screens) is

$$C = \int_0^t \lambda_1 e^{-\lambda_1 s} (1 - e^{-\lambda_2(t-s)}) ds + \frac{I(1 - S)(1 - e^{-\lambda_2 t})}{S}$$

which solves to

$$C = 1 - \frac{(\lambda_2 e^{-\lambda_1 t} - \lambda_1 e^{-\lambda_2 t})}{(\lambda_2 - \lambda_1)} + \frac{I(1 - S)(1 - e^{-\lambda_2 t})}{S}$$

Given these probabilities based on the three formulae for P , I , and C , specific formulae for the log-likelihoods, separately for first and subsequent screens, were specified. After substituting the closed form estimates of S as above, maximum likelihood estimation (MLE) on the total log-likelihood was performed to obtain point estimates for λ_1 and λ_2 . The point estimates of λ_1 and λ_2 were used to calculate the expected values of P , I , and C . The 95% confidence intervals around these expected values were approximated and calculated by the following formula

$$\text{Expected value (e.g. } P) \pm 1.96 \sqrt{\frac{\hat{P}(1 - \hat{P})}{\text{total prevalent screens}}}$$

Method 2: based on non-linear regression

If test sensitivity, calculated under the assumption that interval cancers in the first year after screening are missed cancers, is not adequately describing the real test sensitivity of mammography, a more complex estimation method is needed to estimate the three parameters at once. Then, a three-state Markov model can be applied to depict the progression process of breast cancer from the states of free from breast cancer (state 0), preclinical disease (state 1), and clinical disease (state 2). Breast cancers detected at screens were those in the preclinical detectable phase (PCDP), and interval breast cancers were in the clinical phase (CP).^{24,25} Let the underlying incidence of preclinical disease and the rate of disease progression from preclinical to clinical phase be denoted as above by λ_1 and λ_2 , respectively. The intensity matrix of the three-state model is thus

$$\begin{array}{ccc} & \text{State 0} & \text{State 1} & \text{State 2} \\ \text{State 0} & \left[\begin{array}{ccc} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 1 \end{array} \right] \\ \text{State 1} & & & \\ \text{State 2} & & & \end{array}$$

and the corresponding transition probability matrix for interval between screens in years t

$$\begin{array}{ccc} & \text{State 0} & \text{State 1} & \text{State 2} \\ \text{State 0} & \left[\begin{array}{ccc} P_{00}(t) & P_{01}(t) & P_{02}(t) \\ 0 & P_{11}(t) & P_{12}(t) \\ 0 & 0 & 1 \end{array} \right] \\ \text{State 1} & & & \\ \text{State 2} & & & \end{array}$$

With the following definitions

$$\begin{aligned} P_{00}(t) &= e^{-\lambda_1 t} \\ P_{01}(t) &= \frac{\lambda_1(e^{-\lambda_1 t} - e^{-\lambda_2 t})}{(\lambda_2 - \lambda_1)} \\ P_{02}(t) &= 1 - \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} + \frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} \\ P_{11}(t) &= e^{-\lambda_2 t} \\ P_{12}(t) &= 1 - e^{-\lambda_2 t} \end{aligned}$$

The probabilities of observing preclinical cancers and subjects free from breast cancer in the prevalent screening round are thus

$$\begin{aligned} PCDP &= \left(\frac{P_{01}(v_m)}{P_{00}(v_m) + P_{01}(v_m)} \right) \\ \text{Free from breast cancer} &= \left(\frac{P_{00}(v_m)}{P_{00}(v_m) + P_{01}(v_m)} \right) \end{aligned} \quad (\text{A})$$

where v_m is age at prevalent screen. The probabilities of observing clinical disease, preclinical disease, and

subjects free from breast cancer in the subsequent round are thus

$$\begin{aligned} CP &= P_{01}(t) \times \lambda_2 \\ PCDP &= P_{01}(t) \\ \text{Free from breast cancer} &= P_{00}(t) \end{aligned} \quad (\text{B})$$

where t is the interval between screens in years. For those with preclinical disease or free of breast cancer in the subsequent screen, $t = 2$ due to the biennial screening interval in the Nijmegen breast cancer screening program. For clinical cancers, t is the corresponding time between two screens. Estimation of parameters was performed from the above series of equations by letting the observed numbers equal the expected plus an error term for each mode of detection using non-linear regression (NLR).^{24,25} For the estimation of test sensitivity, the probabilities in formula (A) and (B) were extended as follows

$$\begin{aligned} PCDP &= \left(\frac{P_{01}(v_m)}{P_{00}(v_m) + P_{01}(v_m)} \right) \times S \\ \text{Free from breast cancer} &= \left(\frac{P_{00}(v_m)}{P_{00}(v_m) + P_{01}(v_m)} \right) \\ &+ \left(\frac{P_{01}(v_m)}{P_{00}(v_m) + P_{01}(v_m)} \right) \times (1 - S) \end{aligned} \quad (\text{C})$$

and

$$\begin{aligned} CP &= P_{01}(t) \times \lambda_2 + \frac{F_0}{n_0} \times \frac{1 - S}{S} \times P_{11}(t) \times \lambda_2 \\ PCDP &= P_{01}(t) + \frac{F_0}{n_0} \times \frac{1 - S}{S} \times P_{11}(t) \\ \text{Free from breast cancer} &= P_{00}(t) + (1 - S) \times P_{01}(t) \end{aligned} \quad (\text{D})$$

where F_0 represents the total number of breast cancers detected in the previous screen round and n_0 represents the number of attendees at previous screens.²⁵ Considering the model taking into account the heterogeneity in transition rates, a regression model was adopted as follows

$$\begin{aligned} \lambda_1 &= \lambda_{10} \exp(\beta_1 X) \\ \lambda_2 &= \lambda_{20} \exp(\beta_2 X) \end{aligned}$$

where X is a vector of dummy variables corresponding to each period with the effect on underlying breast cancer incidence rate and rate of progression through regression coefficients β_1 and β_2 , respectively, and λ_{10} and λ_{20} represent the baseline transition rates. The heterogeneity in sensitivity (S) of each period was modeled using a logistic form

$$S = \frac{\exp(rX)}{1 + \exp(rX)}$$

The parameters were estimated by NLR, with the expected numbers of cancers at first and subsequent screens, and numbers of interval cancers based on the above formulae as the regression predictor and the observed numbers as the dependent variable.²⁴

Method 3: based on Markov Chain Monte Carlo simulation

We also applied Markov Chain Monte Carlo simulation (MCMC) to estimate the parameters from the model described above. The same formulae for probabilities and expectations were used, but the parameters were assigned vague prior distributions and estimated in a Bayesian framework using MCMC.²⁶ A Gibbs sampler was used to derive samples of a stationary posterior distribution by which inferences on parameters were drawn. A thinning interval of 3 with a burn-in interval of 10,000 and a total of 15,000 iterations were used, which yields a total of 5000 updated posterior samples.

Results

The numbers of invited, screened, and referred women over the four periods, as well as the numbers of screen-detected and interval cancers, are presented in Table 1. The attendance rate steadily increased over the four periods. The referral rate was stable in periods 1 and 2, but increased during the last two periods, to 29 per 1000 women screened. The first rise in the referral rate coincided with the optimization study and the recommendation to increase referral.¹⁷ The second rise was seen after the introduction of digital mammography. With digital mammography (period 4), the screen-detected cancer rate increased from 5 to almost 7 tumours per 1000 women screened.^{19,27} The number of screen-detected ductal carcinomas in situ was also higher, however, the interval cancer rate remained stable over the four periods.

Table 2 shows the estimates for test sensitivity, the underlying breast cancer incidence, and mean sojourn time over the four periods calculated by each of the three methods (MLE, NLR, MCMC). The estimated underlying breast cancer incidence has risen with approximately 1 case per 1000 women per year in the final period compared with the first for all methods. The calculated test sensitivity of mammography using MLE was around 90%. In contrast, the estimated test sensitivity using NLR and MCMC was higher in period 4 (86% and 79%, respectively) compared with the previous periods. The estimates of mean sojourn time remained fairly stable over the first three periods, whereas the point estimates of all methods showed a small increase in period 4 compared with period 3 (MLE: 2.4 vs. 3.3, NLR: 3.6 vs. 4.4, MCMC: 4.3 vs. 4.6). The mean sojourn time estimated based on MLE was three years, while the estimates of NLR and MCMC were closer to four years.

Table 1. Number of invited, screened, and referred women (aged 50–69) and number of screen-detected and interval-detected cancers for all screens and first and subsequent screens separately in the Nijmegen Breast Cancer Screening Program in the period 1975–2012.

	Period 1 1975–1988	Period 2 1989–2000	Period 3 2001–2006	Period 4 2007–2012
<i>All screens</i>				
Invited	99,702	85,320	46,816	51,767
Screened	65,404	59,208	35,958	40,107
Attendance rate (%)	66	69	77	78
Referred	545	493	504	1168
Referral rate (per 1000 women screened)	8.3	8.3	14.0	29.1
Screen-detected cancer ^a	260 (40)	263 (35)	175 (24)	263 (60)
Screen-detected rate (per 1000 women screened)	4.0	4.4	4.9	6.6
Interval cancer	130	121	82	84
Interval cancer rate (per 1000 women screened)	2.0	2.0	2.3	2.1
Interval cancer year 1	38	38	20	30
Interval cancer year 2	92	83	62	54
<i>First screens</i>				
Screened	12,317	3,007	4,163	4,907
Attendance rate (%)	96.4	90.7	96.9	88.6
Referred	182	50	100	306
Referral rate (per 1000 women screened)	14.8	16.6	24.0	62.4
Screen-detected cancer	69	23	20	49
Screen-detected rate (per 1000 women screened)	5.6	7.6	4.8	10.0
Interval cancer	18	7	11	13
Interval cancer year 1	3	3	3	5
Interval cancer year 2	15	4	8	8
<i>Subsequent screens</i>				
Screened	53,087	56,201	31,795	35,200
Attendance rate (%)	62.6	68.5	74.8	76.1
Referred	363	443	404	862
Referral rate (per 1000 women screened)	6.8	7.9	12.7	24.5
Screen-detected cancer	191	240	155	214
Screen-detected rate (per 1000 women screened)	3.6	4.3	4.9	6.1
Interval cancer	112	114	71	71
Interval cancer year 1	35	35	17	25
Interval cancer year 2	77	79	54	46

^aIn parentheses, the number of ductal carcinoma in situ (DCIS) out of the total number of cancers.

Discussion

Our analysis over 40 years of breast cancer screening in Nijmegen showed a trend towards higher test sensitivity and longer mean sojourn time after the introduction of digital mammography. The three methods used for obtaining estimates of mammography test sensitivity and mean sojourn time gave comparable results. However, the more sophisticated methods, NLR and MCMC, which also estimated test sensitivity from empirical screening data, showed more realistic estimates.

The three methods applied in this study have been used previously to estimate test sensitivity and mean sojourn time, but have never been used to investigate these parameters over time as well as been directly compared with one another before. Our estimates of mean sojourn time during the period of analogue mammography (period 1–3) are comparable with previous results.^{24,25} In contrast, the results of period 4 are difficult to compare with what is known, because to our knowledge there are no other

estimates of mean sojourn time based on digital mammography. The slight increase in test sensitivity after the introduction of digital mammography has been found previously.^{19,27,28}

A major strength of our study is that we were able to investigate the consequences of changes in the screening program on important parameters for the screening interval over a 40-year follow-up period in a well-documented screening program. As the Nijmegen screening program has been running for decades, there were no women who were not invited for screening or reliable estimate of the underlying breast cancer incidence available. Therefore, the underlying breast cancer incidence also needed to be estimated. The estimated underlying breast cancer incidence showed an upwards trend over the four periods. A similar trend in the breast cancer incidence was seen in women aged 45–49 in the Netherlands (unscreened population).²⁹ This can be explained by opportunistic screening and an increase in breast cancer risk factors.

Table 2. Estimation of mammography sensitivity (S), mean sojourn time (MST), underlying breast cancer incidence (λ_1), using three estimation methods applied to data from women aged 50–69 invited to participate in the Nijmegen Breast Cancer Screening Program in the period 1975–2012.

	Period 1 1975–1988			Period 2 1989–2000			Period 3 2001–2006			Period 4 2007–2012		
	S (%)	MST (years)	λ_1 (per 1000 women per year)	S (%)	MST (years)	λ_1 (per 1000 women per year)	S (%)	MST (years)	λ_1 (per 1000 women per year)	S (%)	MST (years)	λ_1 (per 1,000 women per year)
Method 1 MLE	87 (83–91)	2.5 (2.1–2.9)	2.8 (2.5–3.0)	87 (84–91)	2.9 (2.4–3.5)	3.2 (2.9–3.4)	90 (85–94)	2.4 (2.0–3.0)	3.4 (3.1–3.8)	90 (86–93)	3.3 (2.8–4.0)	4.0 (3.6–4.4)
Method 2 NLR	81 (60–92)	3.1 (2.2–5.2)	2.0 (1.7–2.3)	80 (35–97)	3.0 (1.4–6.6)	2.9 (1.7–5.0)	72 (24–95)	3.6 (1.6–8.2)	2.4 (1.4–3.9)	86 (43–98)	4.4 (1.9–10.0)	3.0 (1.8–5.0)
Method 3 MCMC	75 (65–84)	3.4 (2.7–4.3)	2.4 (2.1–2.6)	75 (67–81)	3.4 (2.9–4.2)	2.5 (2.3–3.6)	70 (58–82)	4.3 (3.1–5.7)	3.2 (2.8–3.6)	79 (70–87)	4.6 (3.7–5.9)	3.5 (3.2–3.9)

MLE: maximum likelihood estimation; NLR: non-linear regression; MCMC: Markov Chain Monte Carlo simulation. Within parentheses, 95% confidence interval.

Furthermore, the stable interval cancer rates and higher screen detection rates support a real increase in the incidence, and not an increase in overdiagnosis. We also found that the results of the three methods correlated well, which may suggest that the estimates are accurate, also in comparison with previously published results for screening with analogue mammography.^{24,30}

Another strength is that our approach can easily be applied to other (breast) cancer screening programs to investigate the impact of changes in a screening program on test sensitivity and sojourn time. The MLE method, with a very simple closed form approximation for sensitivity, gave rather high estimates for sensitivity in the early periods, but is fairly easy to use. Moreover, this method is more constrained than the other two methods and arguably gives confidence intervals that are too small, not reflecting the entire uncertainty in the data. It may be that the approximations used in this method were less accurate than those in the more sophisticated methods (NLR, MCMC). NLR assumes a Poisson distribution of observed total numbers with their expectations, which leads to very wide confidence intervals. MCMC takes account of the conditionalities of these distributions on the distributions of related variables, and could, therefore, be argued to model more closely the interrelationships in the data and the model parameters. Thus, MCMC has arguably the most realistic confidence interval estimates. As there is no gold standard for estimating mean sojourn time, it is difficult to judge which of these three estimation methods has the most reliable outcome. Furthermore, we chose methods that can be applied to routinely collected screening data instead of more complex data, such as tumour size for calculating tumour volume/doubling times.

An important limitation is the size of our dataset, as we use data from a breast cancer screening program in a single city in the Netherlands. Because of the limited number of cancers detected and wide confidence intervals, our results need to be interpreted with caution. Nonetheless, all three methods, especially NLR and MCMC, gave the impression that mean sojourn time was longer after the introduction of digital mammography. The point estimate of mean sojourn time is almost 1.5 years longer than when the screening program started (NLR: 3.1 vs. 4.4 year, MCMC: 3.4 vs. 4.6 year). This may be seen as the first piece of evidence towards considering a longer screening interval. To be more certain about our findings, validation of the results in a larger dataset would be the next step. Furthermore, as not all breast cancers have a mammographically detectable preclinical phase, we can estimate the mean sojourn time only as an average for the breast cancers that can be detected by mammography and those that cannot. Moreover, it is likely that the prognosis of the cancer in these two groups is different, which may also have affected the estimates of mean sojourn time.³¹ The potential introduction of new or additional screening modalities, such as digital breast

tomosynthesis, automated breast ultrasound, and breast MRI, will probably make it possible to estimate the mean sojourn time for a larger proportion of all breast cancers. Future research should also include estimating test sensitivity and mean sojourn time for subgroups of women at varying levels of breast cancer risk. This could underpin the length of the screening interval for subgroups of women in the move towards a risk-based breast cancer screening program, rather than the one-size-fits-all approach based on age alone.

Conclusion

This study shows that test sensitivity and mean sojourn time, while taking the underlying breast cancer incidence rate into account, can be investigated based on routinely available screening data, with more complex methods providing the most realistic outcomes.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Netherlands HCot. *Population screening for breast cancer: expectation and developments*. The Hague: Health Council of the Netherlands, 2014.
2. Verbeek AL, Hendriks JH, Holland R, et al. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet* 1984; 1: 1222-1224.
3. Collette HJ, Day NE, Rombach JJ, et al. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *Lancet* 1984; 1: 1224-1226.
4. Prevost TC, Launoy G, Duffy SW, et al. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *Am J Epidemiol* 1998; 148: 609-619.
5. Jiang H, Walter SD, Brown PE, et al. Estimation of screening sensitivity and sojourn time from an organized screening program. *Cancer Epidemiol* 2016; 44: 178-185.
6. Lundgren B. Observations on growth rate of breast carcinomas and its possible implications for lead time. *Cancer* 1977; 40: 1722-1725.
7. Otten JD, van Schoor G, Peer PG, et al. Growth rate of invasive ductal carcinomas from a screened 50-74-year-old population. *J Med Screen* 2018; 25: 40-46.
8. Zelen M and Feinleib M. On the theory of screening for chronic diseases. *Biometrika* 1969; 56: 601-614.
9. Day NE and Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984; 40: 1-14.
10. Paci E and Duffy SW. Modelling the analysis of breast cancer screening programmes: sensitivity, lead time and predictive value in the Florence District Programme (1975-1986). *Int J Epidemiol* 1991; 20: 852-858.
11. Chen JS and Prorok PC. Lead time estimation in a controlled screening program. *Am J Epidemiol* 1983; 118: 740-751.
12. van Oortmarssen GJ, Habbema JD, Lubbe JT, et al. A model-based analysis of the HIP project for breast cancer screening. *Int J Cancer* 1990; 46: 207-213.
13. Duffy SW, Chen HH, Tabar L, et al. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat Med* 1995; 14: 1531-1543.
14. van Luijt PA, Fracheboud J, Heijnsdijk EA, et al. Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens. *Eur J Cancer* 2013; 49: 3517-3525.
15. Otten JD, van Dijck JA, Peer PG, et al. Long term breast cancer screening in Nijmegen, The Netherlands: the nine rounds from 1975-92. *J Epidemiol Community Health* 1996; 50: 353-358.
16. Gezondheidsraad. *De vroeg opsporing van borstkanker*. Den Haag: Gezondheidsraad, 1987.
17. Otten JD, Karssemeijer N, Hendriks JH, et al. Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators. *J Natl Cancer Inst* 2005; 97: 748-754.
18. Timmers JM, den Heeten GJ, Adang EM, et al. Dutch digital breast cancer screening: implications for breast cancer care. *Eur J Public Health* 2012; 22: 925-929.
19. Bluekens AM, Holland R, Karssemeijer N, et al. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. *Radiology* 2012; 265: 707-714.
20. Michell M. *Breast cancer*. London: Cambridge University Press, 2010
21. Michalopoulos D and Duffy SW. Estimation of overdiagnosis using short-term trends and lead time estimates uncontaminated by overdiagnosed cases: results from the Norwegian Breast Screening Programme. *J Med Screen* 2016; 23: 192-202.
22. Tabar L, Fagerberg G, Duffy SW, et al. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992; 30: 187-210.
23. Duffy SW, Agbaje O, Tabar L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005; 7: 258-265.
24. Chen HH, Duffy SW and Tabar L. A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity, and positive predictive value for mammography in breast cancer screening. *Stat* 1996; 45: 307-317.
25. Chen TH, Kuo HS, Yen MF, et al. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* 2000; 56: 167-172.
26. Chien CR, Lai MS and Chen TH. Estimation of mean sojourn time for lung cancer by chest X-ray screening with a Bayesian approach. *Lung Cancer* 2008; 62: 215-220.
27. Bluekens AM, Karssemeijer N, Beijerinck D, et al. Consequences of digital mammography in population-based breast cancer screening: initial changes and long-term impact on referral rates. *Eur Radiol* 2010; 20: 2067-2073.
28. Karssemeijer N, Bluekens AM, Beijerinck D, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 2009; 253: 353-358.
29. Netherlands Ik. *Cijfers over kanker 2018*, https://www.cijfersoverkanker.nl/selecties/dataset_1/img5c04f5ba05a7b
30. Shen Y and Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *JCO* 2001; 19: 3490-3499.
31. Tabar L, Dean PB, Yen AM, et al. A proposal to unify the classification of breast and prostate cancers based on the anatomic site of cancer origin and on long-term patient outcome. *Breast Cancer* 2014; 8: 15-38.