



Original article

Mammographic sensitivity as a function of tumor size: A novel estimation based on population-based screening data



Jing Wang^{a,1}, Pam Gottschal^{a,1}, Lili Ding^a, DaniëlleW.A van Veldhuizen^a, Wenli Lu^b, Nehmat Houssami^c, Marcel J.W. Greuter^{d,e}, Geertruida H. de Bock^{a,*}

^a University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands

^b Department of Epidemiology and Health Statistics, School of Public Health, Tianjin Medical University, Tianjin, China

^c Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Australia

^d University of Groningen, University Medical Center Groningen, Department of Radiology, Groningen, the Netherlands

^e Robotics and Mechatronics (RaM) Group, Faculty of Electrical Engineering Mathematics and Computer Science, Technical Medical Centre, University of Twente, Enschede, the Netherlands

ARTICLE INFO

Article history:

Received 23 September 2020

Received in revised form

24 November 2020

Accepted 7 December 2020

Available online 9 December 2020

Keywords:

Breast Neoplasms

Mass screening

Mammography

Sensitivity

Tumor growth

ABSTRACT

Background: Instead of a single value for mammographic sensitivity, a sensitivity function based on tumor size more realistically reflects mammography's detection capability. Because previous models may have overestimated size-specific sensitivity, we aimed to provide a novel approach to improve sensitivity estimation as a function of tumor size.

Methods: Using aggregated data on interval and screen-detected cancers, observed tumor sizes were back-calculated to the time of screening using an exponential tumor growth model and a follow-up time of 4 years. From the observed number of detected cancers and an estimation of the number of false-negative cancers, a model for the sensitivity as a function of tumor size was determined. A univariate sensitivity analysis was conducted by varying follow-up time and tumor volume doubling time (TVDT). A systematic review was conducted for external validation of the sensitivity model.

Results: Aggregated data of 22,915 screen-detected and 10,670 interval breast cancers from the Dutch screening program were used. The model showed that sensitivity increased from 0 to 85% for tumor sizes from 2 to 20 mm. When TVDT was set at the upper and lower limits of the confidence interval, sensitivity for a 20-mm tumor was 74% and 93%, respectively. The estimated sensitivity gave comparable estimates to those from two of three studies identified by our systematic review.

Conclusion: Derived from aggregated breast screening outcomes data, our model's estimation of sensitivity as a function of tumor size may provide a better representation of data observed in screening programs than other models.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is the most common cancer and one of the main causes of death in European women, approximately one in seven

women will develop breast cancer by the age of 75 [1]. In recent decennia, mammography screening has been introduced in many countries. Studies have shown that screening can detect breast cancer at an earlier stage which will reduce treatment burden and improve survival [2–4]. However, there are ongoing debates on whether screening does more harm than good and on the related optimization of screening strategy. To inform these debates, it is important to evaluate breast cancer screening programs considering indicators of both long-term, such as decreasing burden of breast cancer-specific treatment and mortality benefit, and short-term indicators such as mammography sensitivity and specificity [5,6]. In this contribution, we focus on the estimation of mammographic sensitivity as a function of tumor size, which is highly

Abbreviations: FN, false negative; TP, true positive; TVDT, tumor volume doubling time.

* Corresponding author.

E-mail addresses: j.wang@umcg.nl (J. Wang), p.gottschal.1@student.rug.nl (P. Gottschal), l.ding@umcg.nl (L. Ding), w.a.van.veldhuizen@student.rug.nl (DaniëlleW.A Veldhuizen), luwenli@tmu.edu.cn (W. Lu), nehmat.houssami@sydney.edu.au (N. Houssami), m.j.w.greuter@umcg.nl (M.J.W. Greuter), g.h.de.bock@umcg.nl (G.H. de Bock).

¹ Both authors contributed equally to this work.

relevant for the evaluation of screening programs [7,8]. However, we cannot measure sensitivity directly as there are no methods to determine the amount of asymptomatic cancers that are detectable by screening [9].

Whereas most studies give one constant estimate for the sensitivity of mammography, Weedon-Fekjær et al. developed a logistic model to estimate the sensitivity of mammography as a function of tumor size [10,11]. In their studies, the sensitivity was estimated simultaneously with a continuous growth model utilizing breast cancer screening data, and back-calculation methods were used to estimate tumor size at screening from tumor size distributions of clinically detected tumors. Inspired by this approach, Swedish researchers estimated the sensitivity not only based on tumor size, but also breast density [12,13]. What is remarkable about the findings of their studies is that the sensitivity is 100% for tumors varying in size from 15 to 20 mm and over. However, this seems unlikely, as several studies showed that approximately 10–30% of all screen-detected tumors are larger than 20 mm, which indicated that at least a part of these tumors are missed at the size of 15–20 mm [14–16]. In addition, studies have shown that even tumors larger than 50 mm can be invisible on mammography [17,18].

In this study, we therefore aimed to provide a novel method to improve the estimates of mammography sensitivity as a function of tumor size by using aggregated data reported from a national population breast cancer screening program. We anticipate that the sensitivity function can be integrated into modeling studies focusing on the evaluation of breast cancer screening programs, which in turn can provide valuable evidence for the optimization of screening strategies.

2. Methods

A sensitivity model estimating mammographic sensitivity as a function of tumor size was developed in this study. To develop this sensitivity model, empiric data on number and sizes of screen-detected and interval cancers from a population-based breast screening program and back-calculation of these tumor sizes to the screening moment were used to determine the number of false negatives (FN). The model was externally validated on published data identified by a systematic review.

2.1. The sensitivity model: a description

In our sensitivity model, the probability of finding a tumor with volume V at screening moment i is based on the well-known formula for sensitivity:

$$S_i(V) = \frac{TP_i(V)}{TP_i(V) + FN_i(V)}$$

where $S_i(V)$ is the sensitivity to detect a tumor of volume V at screening round i , and $TP_i(V)$ and $FN_i(V)$ are the number of true-positives and false-negatives at screening round i as a function of tumor volume V respectively.

To determine the number of false-negatives as a function of volume we use the assumption that the undetected tumors at screening round i grow larger over time and will eventually be detected either at a subsequent screening round or as an interval cancer (Fig. 1).

Let the number of screen-detected tumors at screening moment t_i be equal to N_i , where $i = 1, 2, \dots, M$ runs over the total number M of screening moments in the screening program. Let the size of a tumor k which is screen-detected at screening moment t_i be equal to V_{ki} . Let the number of interval tumors between screening moment i

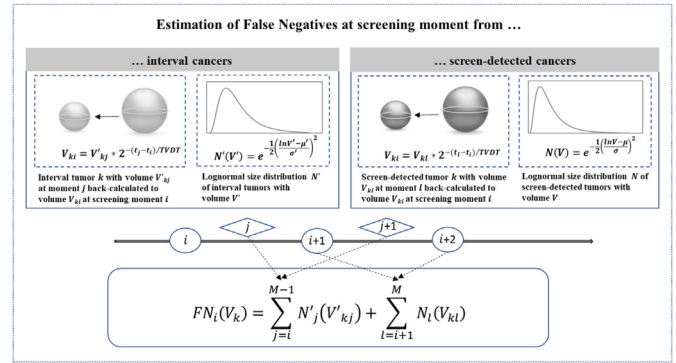


Fig. 1. Estimation of false-negatives (FNs), where V_{ki} represents the volume of a tumor k at t_i , V'_{kj} and V_{kl} represent the volume of a tumor k during screening intervals ($j, j + 1, \dots$) and at subsequent screening rounds ($i + 1, i + 2, \dots$) respectively, and the corresponding numbers of tumors are represented as N'_j and N_l . TVDT = Tumor volume doubling time.

and $i+1$ be equal to N'_j , where $j = 1 \dots M-1$ with corresponding tumor sizes V'_{kj} . Assume for each tumor k an exponential growth model where the volume at screening moment i is given by: $V_{ki} = V_k^0 * 2^{-(t_i - t_0)/TVDT}$, where V_k^0 is the starting volume at time t_0 and $TVDT$ is the tumor volume doubling time. Now, from the tumor size detected by screening or intervals later than screening moment i , we can calculate back the tumor size at the time of screening using the exponential growth model.

If we assume an interval tumor is found at time t_j , then the size of this tumor at screening moment t_i will be equal to: $V_{ki} = V'_{kj} * 2^{-(t_i - t_j)/TVDT}$. Also, the size of a tumor found in one of the subsequent screening rounds $l > i$ can be calculated back in time to the size at the time of screening. If we assume a screen-detected tumor is found at time t_l then the size at screening moment t_i will be equal to: $V_{ki} = V_{kl} * 2^{-(t_i - t_l)/TVDT}$. Now, we can estimate the number of false negatives $FN_i(V)$ with volume V at screening moment i by

$$FN_i(V) = \sum_{j=i}^{M-1} N'_j(V'_{kj}) + \sum_{l=i+1}^M N_l(V_{kl})$$

i.e. the number of back-calculated interval tumors (N'_j) with size V at the time of screening plus the number of back-calculated subsequent screen-detected tumors (N_l) with size V at the time of screening t_i . Together with the number of detected tumors at screening moment i given by $TP_i(V) = N_i(V)$, we can calculate the sensitivity $S_i(V)$ as a function of volume V .

2.2. The sensitivity model: input parameters

2.3. Tumor growth

For tumor growth, tumors were assumed to be spherical and to grow exponentially with a constant volume doubling time [19]. In this study, the tumor volume doubling time for women aged 50–70 years old was on average 157 days [20]. For the distribution of the screen-detected and interval tumor sizes, we used data from the Dutch breast cancer screening program from 2004 to 2009 [14, Table 1]. The data from the first screening round was excluded as it is well known that in the first screening round relatively more and larger tumors are found compared to the subsequent screening rounds [21]. We used a nonlinear least-squares method to obtain the parameters of the log-normal tumor size distributions of the screen-detected and interval cancers found in the screening.

Table 1
Tumor size distribution of screen-detected cancers and interval cancers^a.

T categories (tumor size)	Distribution of screen-detected cancers	Distribution of interval cancers
T1a (≤ 5 mm)	7.7%	2.2%
T1b (>5 mm and ≤ 10 mm)	24.5%	8.9%
T1c (>10 mm and ≤ 20 mm)	49.0%	40.0%
T2 (>20 mm and ≤ 50 mm)	17.9%	41.1%
T3 (>50 mm)	0.9%	7.8%

^a Data source: National evaluation of breast cancer screening in the Netherlands, 1990–2011/2012.

2.4. Time since previous screening

We assumed biennial screening frequency as used in the Dutch screening program and many population-based breast screening programs. The maximum delay time in diagnosis after a false negative breast assessment in recalled women in a biennial screening program was 1251 days, which was rounded up to four years [22]. A median time from biennial screening to diagnosis of interval cancers of 502 days was used [23]. The time between the diagnosis of an interval cancer which had a possible false-negative result in the previous one or two screenings rounds was therefore set at 502 and 1232 (two years plus 502 days) days respectively. The time between the diagnosis of a screen-detected cancer which had a possible false-negative result in the previous one or two screening rounds was set at two years and four years, respectively.

2.5. Analysis of the results of the sensitivity model

The main outcome, i.e. tumor size-specific sensitivity estimated from the developed model was described graphically. To evaluate the uncertainty of our model, univariate sensitivity analyses were performed by varying input values of model parameters. Lastly, external validation of the developed model was conducted based on published data identified by a systematic review.

2.6. Analysis on the assumptions of the sensitivity model

We performed sensitivity analyses to evaluate the uncertainty of our model. The tumor volume doubling time was set to the lower and upper bounds of the 95% confidence interval (CI), which were 121 days and 204 days respectively [20], and the follow-up time between a screen-detected or interval cancer and the previous screening rounds was set at 2 and 6 years.

2.7. External validation of the sensitivity model

For external validation, we performed a systematic search in PubMed to find related articles focusing on mammography sensitivity and tumor size. The keywords used in the search included “breast carcinoma”, “mammography”, “sensitivity and specificity”, and “tumor size”. If the study reported observed sensitivities and related tumor size from a population-based screening program, then it would be included for further comparison. To ensure recent mammographic methods were used, the searches focused on relevant articles published from January 1, 2000 to August 1, 2020. Two authors searched the literature independently. A detailed description of the search strategies can be found in Supplementary data. From the included studies, the reported sensitivity was compared to our model.

3. Results

3.1. Mammography sensitivity according to the model

The aggregated data of 22,915 screen-detected cancers and 10,670 interval cancers were used for the estimation of tumor size distributions. For screen-detected cancers, the mean diameter and corresponding standard deviation (mm) were 14.0 (95%CI:

10.6–18.4) and 1.93 (95%CI: 1.52–2.46), while for interval cancers, these were 20.9 (95%CI: 18.5–23.8) and 1.77 (95%CI: 1.58–1.95), respectively. Given a TVDT of 157 days and a 4-year follow-up, the model showed a sensitivity function which continuously increased from 0 to 85% for tumor diameters between 2 and 20 mm (Fig. 2: Solid line). The estimated sensitivity at 5, 10, 15 and 50 mm was 35%, 65%, 78%, and 97%, respectively.

3.2. Analysis on the assumptions

The mammography sensitivity increased with a decrease in TVDT (Fig. 2). When the TVDT was set at the upper and lower limits of its confidence interval, the sensitivity for a 20-mm tumor became 74% and 93%, respectively. Unlike TVDT, different follow-up times only had a minor impact on our sensitivity model. Increasing the follow-up time to 6 years did not affect the sensitivity. With a shorter follow-up time (2 years), the sensitivity as found with our sensitivity model slightly increased when tumor size was smaller than 10 mm. Specifically, the sensitivity was 39% and 67% for a 5 and a 10 mm tumor, respectively, whereas for larger tumor sizes, the sensitivity remained nearly unchanged.

3.3. External validation of the sensitivity model

After literature searching and screening, three studies were included [24–26]. All three studies reported mammography sensitivity and its related mean tumor size. To allow comparison, our estimated sensitivity at the mean tumor size reported in the literature (Table 2) was used. Specifically, our model gave reliable estimations which were comparable to two of the included studies [25,26]. However, the sensitivity was slightly underestimated compared to that of Cawson et al. [24].

4. Discussion

We developed a novel model for the estimation of mammographic sensitivity as a continuous function of tumor size, given that mammography's detection capability varies according to tumor size. Therefore, such a model provides more details about the sensitivity of mammography screening. Aggregated data of 22,915 screen-detected and 10,670 interval cancers from the Dutch screening program were used to obtain the size distribution of detected as well as missed breast cancers at the time of screening. The estimated sensitivity showed an increase from 0 to 85% for tumors between 2 and 20 mm. A sensitivity analysis for the model indicated that TVDT was an influential factor for sensitivity, and the assumption that a tumor will be detected in a biennial screening program within 2 screening rounds after one false-negative result was deemed appropriate.

In our model, the follow-up time was used to determine how long the expected time was that allows all false-negatives to be detected, so that sensitivity was not overestimated due to underestimations of false negatives. In this study, the follow-up time was estimated based on the maximum delay time in diagnosis

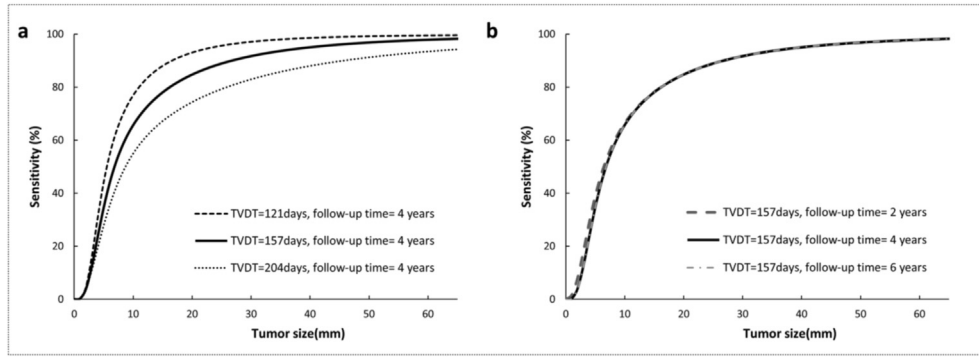


Fig. 2. The estimated mammographic sensitivity as a function of tumor size. Solid line in a and b: the sensitivity model using a TVDT of 157 days and a follow-up time of 4 years; a: the sensitivity analysis of varying TVDT (121 and 204 days), and b: the sensitivity analysis of varying follow-up time (2, and 6 years). TVDT = tumor volume doubling time.

Table 2
Validation results based on screening data.

Reference	Mean tumor size (mm)	Observed sensitivity (%)	Estimated sensitivity at the same size (%)
Cawson et al. [24]	18.7	90.4 (84.7–94.6)	83.6
Moshina et al. [25]	15.6	77.6 (75.6–79.6)	79.2
Skaane et al. [26]	13.6	76.2 (72.2–80.0)	75.7

(1251 days) reported in Ciatto et al., and we used a follow-up time of 4 years which was rounded up from the value 1251 days [22]. Although this study dates from 1992 to 2001, we estimated that this data on follow-up times is still valid in the current state of screening programs. First, according to a recently published study which compared the median delay time between two time periods of 1997–2006 and 2007–2016 in the Netherlands [27], the median delay time for both periods was approximately 2 years which was similar to the reported median delay time in Ciatto et al. Importantly, the difference in delay time between the two periods was not statistically significant. Second, delayed diagnosis after false negative results is not only related to mammography sensitivity itself, but also related to participants compliance as shown in Ciatto et al. [22]. Third, the analysis on assumptions of our model showed that for a shorter follow-up time of 2 years, the screening sensitivity slightly increased, whereas for a longer follow-up time of 6 years, the sensitivity curve barely changed compared to that of 4 years. These results indicate that a follow-up time of 4 years is reasonable.

The validation of our model showed that the estimated sensitivity was comparable to two of the three studies [25,26], whereas the sensitivity was slightly underestimated compared to that of Cawson et al. [24]. A possible reason is that in the study of Cawson et al., only tumors that could be detected or were visible on mammograms were included. However, it is well-known that in a real-world screening setting a proportion of tumors is not detectable by mammography [18], which could explain the higher sensitivity reported. Although the model was generated based on data from Dutch breast cancer screening program, we anticipated that this model could also be applicable globally to other organized population screening programs with biennial mammography like Norway and Australia. This assumption was informed by several studies which suggest that screening interval plays a vital role in estimating mammography sensitivity [28]. Second, in addition to tumor size, mammography sensitivity can also be affected by participants’ characteristics such as mammographic density, and technical factors such as interpretive skills of radiologists [29]. We expect that these factors might not differ much between the Netherlands and the other two countries, and therefore could be used as reliable sources for our external validation.

Compared to other models where a seemingly optimistic

sensitivity of 100% for tumor diameters of 15–20 mm was estimated [10–13], our model provides a more reliable sensitivity of 85% for a tumor diameter of 20 mm. Studies have shown that on average 20% of the screen-detected cancers were larger than 20 mm, and data from Germany showed that approximately 8% of the incident tumors in their population screening program were even larger than 50 mm [14–16]. Although infrequent screening or fast-growing cancers could partly explain these larger tumor sizes, it is unlikely that the sensitivity would reach a perfect sensitivity at a size of 20 mm. In addition, several studies have shown that some cancers will not be visible on mammograms even at a very large size >70 mm [18,30,31], as factors like sites where visualization is difficult (close to the thorax wall), and especially high-density breast tissue will lead to not-detectable tumors on mammograms [18,32].

The shape of the sensitivity curve in our study was similar to models estimated from logistic functions [Figs. 3, 10–13]. However, our model showed a higher sensitivity for tumors ≤10 mm, while the sensitivity became lower when the tumor size was larger than 10 mm. Possible reasons could be mainly explored from the model structure perspective. First, in our model, we did not make a prior

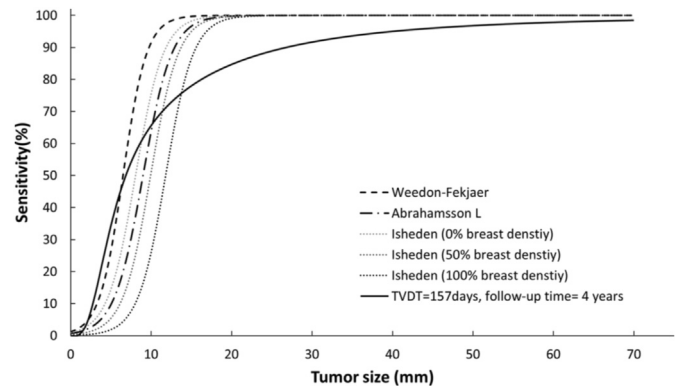


Fig. 3. A comparison of the proposed sensitivity function model with other model studies. Data for 100%, 50%, and 0% breast density of Isheden et al. [12], of Weedon-Fakjaer et al. [10], and of Abrahamsson et al. [13]. TVDT = tumor volume doubling time.

assumption on the sensitivity function itself such as a logistic function that was used in other studies. By assuming a logistic function, the sensitivity would increase sharply at a certain point as observed in Fig. 3, which might lead to a higher sensitivity for larger tumors than that of our model. Second, certain aspects of the tumor growth model might be possible reasons. To be specific, in our model, we assumed that tumor grows through an exponential function with a constant TVDT, according to Collins et al. [19]. However, in Weedon-Fekjær et al., they used a growth model made by Spratt et al., in which tumor was assumed to grow through a logistic function with variation in individual growth rates [33,34]. In studies from Isheden et al. and Abrahamsson et al., although they also used an exponential model, the cell reproductive rate with a constant inverse growth rate was used as their parameters. Tumor growth rate had a crucial impact on the estimation of sensitivity as a faster growing tumor was more likely to be detected at a larger volume than a slower growing tumor, which might result in a higher sensitivity [35]. Nevertheless, it is difficult to compare in a straightforward way as conflicting results were reported on which model performs well and as different parameters were used to express tumor growth in these studies [36,37].

The differences between our model and the other models could be also explored from data perspective. Specifically, to estimate the tumor size distribution, we used relatively new data from 2004 to 2009, while Weedon-Fekjær et al. estimated the parameters of the function by using screening data from 1995 to 2002 [10], and Isheden et al. and Abrahamsson et al. used same data from 1993 to 1995 [12,13]. As mammography detectability has been improved over time [38], and the modern mammography is able to detect more tumors at smaller size, a higher sensitivity at smaller size and also an overall lower sensitivity was observed in our study.

In addition to models that estimate screening sensitivity continuously as a function of tumor size, the MISCAN model calibrated mammographic sensitivity by T-stage. The sensitivity of our model was generally lower than the estimations of sensitivity in studies that used the MISCAN model [39,40]. For example, the estimated sensitivity at ≤ 5 , 5–10, 10–20 and > 20 mm was 47%, 62%, 90% and 98% in Gelder et al. [39], while the estimated sensitivity for a tumor at 5, 10, and 20 mm in our model was 35%, 65%, and 85%. One of the possible reasons could be the inclusion of prevalent cancers detected at the first screening round [40], which might lead to a higher sensitivity in the MISCAN model.

The strengths of this study lie in several aspects. Firstly, unlike other modeling studies that assumed a logistic function [10–13], we estimated a sensitivity model without any prior assumptions. Secondly, we used real-world aggregated data such as the number and size distribution of breast cancers, which can be relatively easily found in the national reports of breast cancer screening programs. Furthermore, the developed model can be easily adapted with different input parameters such as growth rates, different tumor size distributions and interval periods, which could make our sensitivity model useful for screening evaluation in other countries or other screening purposes.

However, there are also some limitations. Firstly, we used population-based data such as the number and size distribution of breast cancers based on T-stage, however, with a more detailed tumor size distribution instead of just the T-stage, the estimation of the lognormal distribution parameters would be more reliable. Secondly, the sensitivity model used a constant tumor diameter doubling time for every tumor, while in reality, the tumor growth varies widely among tumors or even for one tumor at different times [19,41]. Ideally, a comprehensive growth model could be incorporated if more detailed data were available. Thirdly, in our model, we assumed that all false-negative cancers would be detected in the future and became larger over time. However,

studies have shown that some cancers will stop growing and even regress, which might lead to an overestimation of the sensitivity [42]. On the other hand, some fast-growing cancers would be recognized as false-negatives instead of new incident cancers, which might underestimate sensitivity [43]. Moreover, we assumed that the time between two screening rounds and the time between the last screening and an interval cancer was fixed. However, knowledge about these time distributions would enable us to better estimate the distribution of tumor sizes at time of screening. Lastly, in addition to tumor size, breast density and age also has an impact on mammographic sensitivity. Studies have shown that mammography sensitivity decreases in women with dense breasts and younger women [44]. The model presented here gives the sensitivity of mammography as used in screening settings for a population of women with mixed breast density and age. However, the sensitivity as a function of tumor size could in principle also be calculated for women with dense or fatty breasts or for different age groups if specific data on these groups of women were available [12].

5. Conclusion

In this study, we developed a model which estimates the sensitivity of mammography as a function of tumor size without any prior assumptions about the function itself. The sensitivity model showed a similar sensitivity curve shape compared with studies that were estimated from logistic function [10–13], but the estimates in our model had a better representation of data observed in other screening programs. Furthermore, as tumor growth is an influential factor for the estimation of sensitivity, future studies that provide more detailed information on tumor progression, such as tumor doubling times, would help in further refining sensitivity estimates. In summary, our work provides knowledge on the tumor size-specific sensitivity of mammography. Our sensitivity model can be incorporated in cost-effectiveness models aiming to evaluate breast cancer screening programs. A tumor size-specific sensitivity might improve the performance of cost-effectiveness modeling compared with models that use only a single value for mammographic sensitivity.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

The author, J Wang, thanks the support from Chinese scholarship council (CSC) for providing a scholarship for her PhD research. N Houssami receives research support through a National Breast Cancer Foundation (Australia) Breast Cancer Research Leadership Fellowship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at

<https://doi.org/10.1016/j.breast.2020.12.003>.

References

- [1] van Der Waal D, Verbeek AL, Den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in The Netherlands: a changing burden. *Eur J Publ Health* 2015;25:320–4.
- [2] Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast cancer screening programmes across the WHO European region: differences among countries based on national income level. *Int J Environ Res Publ Health* 2017;14:452.
- [3] Shaevitch D, Taghipour S, Miller AB, Montgomery N, Harvey B. Tumor size distribution of invasive breast cancers and the sensitivity of screening methods in the Canadian National Breast Screening Study. *J Canc Res Therapeut* 2017;13:562.
- [4] de Munck L, Fracheboud J, de Bock GH, den Heeten GJ, Siesling S, Broeders MJ. Is the incidence of advanced-stage breast cancer affected by whether women attend a steady-state screening program? *Int J Canc* 2018;143:842–50.
- [5] Mandrik O, Zielonke N, Meheus F, Severens JL, Guha N, Herrero Acosta R, et al. Systematic reviews as a 'lens of evidence': determinants of benefits and harms of breast cancer screening. *Int J Canc* 2019;145:994–1006.
- [6] Abrahamsson L, Humphreys K. A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density. *Stat Methods Med Res* 2016;25:1620–37.
- [7] Centers for Disease Control and Prevention. Quality standards compliance for mammography facilities. *J Am Med Assoc* 1994;272:763.
- [8] Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. *J Med Screen* 2007;14:174–7.
- [9] Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. *AJR Am J Roentgenol* 2002;179:825–31.
- [10] Weedon-Felkjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Res* 2008;10:R41.
- [11] Weedon-Felkjaer H, Tretli S, Aalen OO. Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening. *Stat Methods Med Res* 2010;19:507–27.
- [12] Isheden G, Humphreys K. Modelling breast cancer tumour growth for a stable disease population. *Stat Methods Med Res* 2019;28:681–702.
- [13] Abrahamsson L, Isheden G, Czene K, Humphreys K. Continuous tumour growth models, lead time estimation and length bias in breast cancer screening studies. *Stat Methods Med Res* 2020;29:374–95.
- [14] Fracheboud J, Van Luijt PA, Sankatsing VDV. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland. 2014.
- [15] Simbrich A, Wellmann I, Heidrich J, Heidinger O, Hense HW. Trends in advanced breast cancer incidence rates after implementation of a mammography screening program in a German population. *Cancer Epidemiol* 2016;44:44–51.
- [16] Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian breast cancer screening program. *Eur J Epidemiol* 2007;22:447.
- [17] Rajentheran R, Rao CM, Lim E, Lennard TW. Palpable breast cancer which is mammographically invisible. *Breast* 2001;10:416–20.
- [18] Foxcroft LM, Evans EB, Joshua HK, Hirst C. Breast cancers invisible on mammography. *Aust N Z J Surg* 2000;70:162–7.
- [19] Collins VP, Loeffler RK, Tivey H. Observations on growth rates of human tumors. *Am J Roentgenol* 1956;76:988–1000.
- [20] Peer PG, Van Dijck JA, Verbeek AL, Hendriks JH, Holland R. Age-dependent growth rate of primary breast cancer. *Cancer* 1993;71:3547–51.
- [21] Hofvind S, Sørum R, Thoresen S. Incidence and tumor characteristics of breast cancer diagnosed before and after implementation of a population-based screening-program. *Acta Oncol* 2008;47:225–31.
- [22] Ciatto S, Houssami N, Ambrogetti D, Bonardi R, Collini G, Del Turco MR. Minority report—false negative breast assessment in women recalled for suspicious screening mammography: imaging and pathological features, and associated delay in diagnosis. *Breast Canc Res Treat* 2007;105:37–43.
- [23] Wang H, Bjørnstam N, Bjørndal H, Braaten A, Eriksen L, Skaane P, et al. Interval cancers in the Norwegian breast cancer screening program: frequency, characteristics and use of HRT. *Int J Canc* 2001;94:594–8.
- [24] Cawson JN, Nickson C, Amos A, Hill G, Whan AB, Kavanagh AM. Invasive breast cancers detected by screening mammography: a detailed comparison of computer-aided detection-assisted single reading and double reading. *J Med Imaging Radiat Oncol* 2009;53:442–9.
- [25] Moshina N, Sebuødegård S, Lee CI, Akslen LA, Tsuruda KM, Elmore JG, et al. Automated volumetric analysis of mammographic density in a screening setting: worse outcomes for women with dense breasts. *Radiology* 2018;288:343–52.
- [26] Skaane P, Sebuødegård S, Bandos AI, Gur D, Østerås BH, Gullien R, et al. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. *Breast Canc Res Treat* 2018;169:489–96.
- [27] Lameijer JR, Voogd AC, Pijnappel RM, Setz-Pels W, Broeders MJ, Tjan-Heijnen VC, et al. Delayed breast cancer diagnosis after repeated recall at biennial screening mammography: an observational follow-up study from The Netherlands. *Br J Canc* 2020;123:325–32.
- [28] Domingo L, Hofvind S, Hubbard RA, Román M, Benkeser D, Sala M, et al. Cross-national comparison of screening mammography accuracy measures in US, Norway, and Spain. *Eur Radiol* 2016;26:2520–8.
- [29] Baines CJ, Dayan R. A tangled web: factors likely to affect the efficacy of screening mammography. *J Natl Cancer Inst* 1999;91:833–8.
- [30] Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd NF. Case-control study of factors associated with failure to detect breast cancer by mammography. *J Natl Cancer Inst* 1992;84:781–5.
- [31] Wadhwa A, Sullivan JR, Gonyo MB. Missed breast cancer: what can we learn? *Curr Probl Diagn Radiol* 2016;45:402–19.
- [32] Huynh PT, Jarolimiek AM, Daye S. The false-negative mammogram. *Radiographics* 1998;18:1137–54.
- [33] Spratt JA, von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. *Cancer* 1993;71:2013–9.
- [34] Spratt JA, von Fournier D, Spratt JS, Weber EE. Mammographic assessment of human breast cancer growth and duration. *Cancer* 1993;71:2020–6.
- [35] Plevritis SK, Salzman P, Sigal BM, Glynn PW. A natural history model of stage progression applied to breast cancer. *Stat Med* 2007;26:581–95.
- [36] Talkington A, Durrett R. Estimating tumor growth rates in vivo. *Bull Math Biol* 2015;77:1934–54.
- [37] Sarapata EA, de Pillis LG. A comparison and catalog of intrinsic tumor growth models. *Bull Math Biol* 2014;76:2010–24.
- [38] Weedon-Felkjaer H, Vatten LJ, Aalen OO, Lindqvist B, Tretli S. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *J Med Screen* 2005;12:172–8.
- [39] De Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med* 2011;53:134–40.
- [40] Rijsburger AJ, van Oortmarssen GJ, Boer R, Draisma G, Miller AB, de Koning HJ. Mammography benefit in the Canadian national breast screening study-2: a model evaluation. *Int J Canc* 2004;110:756–62.
- [41] Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* 1995;75:2507–17.
- [42] Zahl PH, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008;168:2311–6.
- [43] Brekelmans CT, van Gorp JM, Peeters PH, Collette HJ. Histopathology and growth rate of interval breast carcinoma: characterization of different subgroups. *Cancer* 1996;78:1220–8.
- [44] Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138:168–75.